COGNTIVE 2015

The Seventh International Conference on Advanced Cognitive Technologies and Applications

ISBN: 978-1-61208-390-2

HOLIDES 2015

The First Workshop on Holistic Human Factors for Adaptive Cooperative Human-Machine Systems

March 22 - 27, 2015

Nice, France

COGNITIVE 2015 Editors

Nikos Makris, Massachusetts General Hospital | Harvard Medical School, USA

Giorgio Bonmassar, Massachusetts General Hospital || Harvard Medical School, USA

Victor Raskin, Purdue University - W. Lafayette, USA

Julia M. Taylor, Purdue University - W. Lafayette, USA

Simona Collina, Università degli Studi Suor Orsola Benincasa - Napoli, Italy

COGNITIVE 2015

Forward

The Seventh International Conference on Advanced Cognitive Technologies and Applications (COGNITIVE 2015), held between March 22-27, 2015 in Nice, France, targeted advanced concepts, solutions and applications of artificial intelligence, knowledge processing, agents, as key-players, and autonomy as manifestation of self-organized entities and systems. The advances in applying ontology and semantics concepts, web-oriented agents, ambient intelligence, and coordination between autonomous entities led to different solutions on knowledge discovery, learning, and social solutions.

The conference had the following tracks:

- Brain information processing and informatics
- Artificial intelligence and cognition
- Agent-based adaptive systems
- Applications

Similar to the previous edition, this event attracted excellent contributions and active participation from all over the world. We were very pleased to receive top quality contributions.

COGNITIVE 2015 also included the following workshop:

- HOLIDES 2015, The First Workshop on Holistic Human Factors for Adaptive Cooperative Human-Machine Systems

We take here the opportunity to warmly thank all the members of the COGNITIVE 2015 technical program committee, as well as the numerous reviewers. The creation of such a high quality conference program would not have been possible without their involvement. We also kindly thank all the authors that dedicated much of their time and effort to contribute to COGNITIVE 2015. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

Also, this event could not have been a reality without the support of many individuals, organizations and sponsors. We also gratefully thank the members of the COGNITIVE 2015 organizing committee for their help in handling the logistics and for their work that made this professional meeting a success.

We hope COGNITIVE 2015 was a successful international forum for the exchange of ideas and results between academia and industry and to promote further progress in the area of

cognitive technologies and applications. We also hope that Nice, France provided a pleasant environment during the conference and everyone saved some time to enjoy the charm of the city.

COGNITIVE 2015 Chairs

COGNITIVE Advisory Chairs

Hermann Kaindl, TU-Wien, Austria
Sugata Sanyal, Tata Consultancy Services, Mumbai, India
Po-Hsun Cheng (鄭伯壎), National Kaohsiung Normal University, Taiwan
Narayanan Kulathuramaiyer, UNIMAS, Malaysia
Susanne Lajoie, McGill University, Canada
Jose Alfredo F. Costa, Universidade Federal do Rio Grande do Norte (UFRN), Brazil
Terry Bosomaier, Charles Sturt University, Australia
Hakim Lounis, UQAM, Canada
Darsana Josyula, Bowie State University; University of Maryland, College Park, USA
Om Prakash Rishi, University of Kota, India

COGNITIVE Industry/Research Chair

Qin Xin, Simula Research Laboratory, Norway
Arnau Espinosa, g.tec medical engineering GmbH, Austria
Knud Thomsen, Paul Scherrer Institute, Switzerland

HOLIDES 2015 Co-Chairs

Sebastian Feuerstack, OFFIS e.V/Carl von Ossietzky University-Oldenburg, Germany
Simona Collina, Università degli Studi Suor Orsola Benincasa - Napoli, Italy

COGNITIVE 2015

Committee

COGNITIVE Advisory Chairs

Hermann Kaindl, TU-Wien, Austria
Sugata Sanyal, Tata Consultancy Services, Mumbai, India
Po-Hsun Cheng (鄭伯壎), National Kaohsiung Normal University, Taiwan
Narayanan Kulathuramaiyer, UNIMAS, Malaysia
Susanne Lajoie, McGill University, Canada
Jose Alfredo F. Costa, Universidade Federal do Rio Grande do Norte (UFRN), Brazil
Terry Bosomaier, Charles Sturt University, Australia
Hakim Lounis, UQAM, Canada
Darsana Josyula, Bowie State University; University of Maryland, College Park, USA
Om Prakash Rishi, University of Kota, India

COGNITIVE Industry/Research Chair

Qin Xin, Simula Research Laboratory, Norway
Arnau Espinosa, g.tec medical engineering GmbH, Austria
Knud Thomsen, Paul Scherrer Institute, Switzerland

COGNITIVE 2015 Technical Program Committee

Siby Abraham, University of Mumbai, India
Witold Abramowicz, Poznan University of Economics, Poland
Thomas Ågotnes, University of Bergen, Norway
Rajendra Akerkar, Western Norway Research Institute, Norway
Zahid Akhtar, University of Udine, Italy
Jesús B. Alonso Hernández, Universidad de Las Palmas de Gran Canaria, Spain
Giner Alor Hernández, Instituto Tecnológico de Orizaba - Veracruz, México
Galit Fuhrmann Alpert, eBay Inc. / Interdisciplinary Center (IDC) Herzliya, Israel
Stanislaw Ambroszkiewicz, Institute of Computer Science - Polish Academy of Sciences, Poland
Ricardo Ron Angevin, Universidad de Malaga, Spain
Alla Anohina-Naumeca, Riga Technical University, Latvia
Ezendu Ariwa, London Metropolitan University, UK
Ilkka Arminen, University of Helsinki, Finland
Piotr Artiemjew, University of Warmia and Mazury, Poland
Rafael E. Banchs, Institute for Infocomm Research, Singapore
Jean-Paul Barthès, Université de Technologie de Compiègne, France

Knud Thomsen, Paul Scherrer Institut, Switzerland
Ingo J. Timm, University of Trier, Germany
Luz Abril Torres Méndez, CINVESTAV Saltillo, Mexico
Bogdan Trawinski , Wroclaw University of Technology, Poland
Blesson Varghese, Dalhousie University, Canada
Shirshu Varma, Indian Institute of Information Technology, India
Seppo Väyrynen, University of Oulu, Finland
Sebastian Ventura Soto, Universidad of Cordoba, Spain
Maria Fatima Q. Vieira, Universidade Federal de Campina Grande (UFCG), Brazil
Jørgen Villadsen, Technical University of Denmark, Denmark
Zuoguan Wang, Rensselaer Polytechnic Institute, USA
Michal Wozniak, Wroclaw University of Technology, Poland
Xin-She Yang, Middlesex University London, UK
Jure Žabkar, University of Ljubljana, Slovenia
Bin Zhou, University of Maryland, USA
Fuzhen Zhuang, Institute of Computing Technology - Chinese Academy of Sciences, China

HOLIDES 2015 Co-Chairs

Sebastian Feuerstack, OFFIS e.V/Carl von Ossietzky University-Oldenburg, Germany
Simona Collina, Università degli Studi Suor Orsola Benincasa - Napoli, Italy

Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission or reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article is does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

Table of Contents

Using Tags to Improve Diversity of Sparse Associative Memories

Stephen Larroque, Ehsan Sedgh Gooya, Vincent Gripon and Dominique Pastor

Electronics Department
Télécom-Bretagne (Institut Mines-Télécom)
Email: firstname.lastname@telecom-bretagne.eu

***Abstract*—Associative memories, a classical model for brain long-term memory, face interferences between old and new memories. Usually, the only remedy is to enlarge the network so as to retain more memories without collisions: this is the network's size–diversity trade-off. We propose a novel way of representing data in these networks to provide another mean to extend diversity without resizing the network. We show from our analysis and simulations that this method is a viable alternative, which can perfectly fit cases where network's size is constrained, such as neuromorphic FPGA boards implementing associative memories.**

***Keywords*—neural coding; associative memory; neural network; information theory; graph theory; sparse coding; clique; computational neuroscience.**

I. Introduction

Studying the inner workings of brain memory has increasingly become a major challenge for modern neuroscience, since memory is likely a fundamental building block for higher cognitive functions, such as language, reasoning, creativity and consciousness [1].

Associative memories are a branch of now classical computational models for brain memory. Contrary to the *von Neumann* computing architecture [2], [3], where memory is indexed by attributing a unique address for each data, an associative memory change the representation of data in a way that allows to recover an entry only using an incomplete or noisy portion of that data. Furthermore, these models emphasize greater biological plausibility by satisfying the metabolic constraints the organic brain has to face [4], [5].

However, since transformed data can overlap in associative memories, they suffer from interference: there is a tradeoff between network's size and data diversity (number of different entries possibly stored) [6].

We propose a novel way of representing data in these networks by adding a pairing meta-information among edges, thus relaxing the above mentioned tradeoff by providing another way to extend the network's data diversity.

For this purpose, we will first introduce briefly a classical model in Section 2, then in Section 3 we will extend this model with the pairing strategy. The specific dynamics of this extended model will then be analyzed in Section 4 and simulated in Section 5. Finally, an opening to biological hypotheses will be offered to the reader in Section 6 and this work will be concluded alongside a description of a few future avenues in Section 7.

II. Classical model

We will extend the *clique neural network*, a neural network based auto-associative memory introduced by Gripon et al. [7]. Since this is an associative memory, messages are stored such that it is possible to retrieve them from noisy or partially erased input.

Formally, we call *message* a finite sequence of characters of length χ over the alphabet $[\ell]$ where $[\ell]$ denotes the set of integers between 0 and ℓ, with 0 being a special symbol representing emptiness (this is a *non-value*), and c the number of significant, nonzero symbols in a message. This empty 0 symbol allows to construct sparse messages, because this character has no explicit representation in the network.

Consider a set $\mathcal{M}$ of sparse messages. To store them, Aliabadi et al. [8] propose to use a neural network with χ parts each containing ℓ units. They index each part from 1 to χ to correspond to each character of a message, and in each part they index each unit from 1 to ℓ to correspond to the possible values for that character. By notation abuse, we will make no distinction between a unit and its associated pair of indices. Then, they define a mapping associating any sparse message $m = (m_1, m_2, \ldots, m_\chi)$ with a subset of units in the network:

$$\mu = f(m) = \{(i, m_i), 1 \leq i \leq \chi, m_i \neq 0\} \,. \qquad (1)$$

Rather than storing a message m, Aliabadi et al. [8] propose to store μ. To do so, they connect together all units in μ, embodying a clique into the neural network and effectively learning in one-shot. This process is depicted in Figure 1. This implies that the network is binary: an edge exists or it doesn't.

Storing a set of messages $\mathcal{M}$ in the network is simply the union of every messages' cliques. Because cliques can be overlapping (by sharing at least two units), this representation of information is lossy [8].

The process of retrieving a previously stored message, also called *decoding*, given a partial and/or erroneous input then consists in iterating two steps [9], as shown in Figure 2.

Different ways of operating these two steps (called *retrieval rules*) have been extensively studied by Aboudib et al. [9]. In this work, we choose to define the score of a unit to be the number of activated units it is connected to, called the *Sum-Of-Sum rule* [8]. We then select the units that reach the

 ISBN: 978-1-61208-390-2

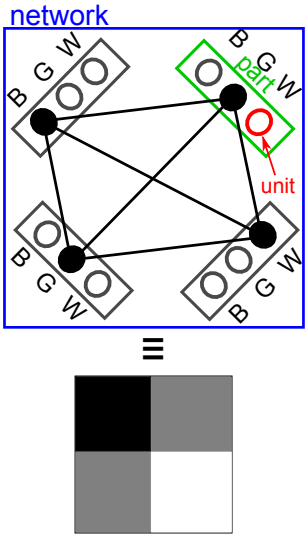


Figure 1. Process of storing an image composed of the pixels sequence {B, G, G, W} (from left to right and from top to down) in the clique network. In this scenario, the parameters of the network are $\chi = c = 4$ and $\ell = 3$ (pixel intensities range of 3 values: B for black, G for gray and W for white).

For each input query:

1) Activate units corresponding to the input query
2) Until stop criterion (number of iterations or convergence criterion):
   a) Propagation: Compute a score s for each unit in the network. This score represents the unnormalized likelihood that a unit is part of the target message.
   b) Filtering: Use a selection operator to choose whether to activate units or not based on their score.

Figure 2. Clique network iterative retrieval (decoding) process

maximum score in the network: this is the *Global Winner-Take-All rule* [10].

III. Proposed Model

Although the clique network is binary, brain synapses do not function in such a fashion: they emit an action potential of variable intensity. Illustrious models for associative memories [11]–[14], as well as most other non-associative neural networks, take account of this variable intensity by affecting a weight on the edges. However, this results in poor performance in terms of memory efficiency, constrained by a sub-linear law [6], [15].

Contrariwise to this approach, we propose to assign a color, or *tag*, to each edge instead of a weight, with the goal of pairing together the edges from the same clique. Indeed, this edge meta-information now represents a pairing cue instead of a synaptic potential intensity modulation. Thus, this meta-information does not affect information processing, but only helps in disambiguating. The tags can be seen as a modified Hebbian rule: *Neurons that fire together, wire together*, and with a strong affinity. In this sense, the tags can be related to the neurobiological mechanism called synaptic discrete states [16]–[18].

More specifically, let us now suppose that connections in the network are not binary but can take up to g distinct, discrete, values. This results in a colored graph, where each connection has its own color. We modify the storage process as follows:

1) First we associate each message to store with a tag,
2) When storing the clique equivalent of a message, we assign the corresponding tag to the clique's edges, replacing any previous tag if the edge already exists.

As a result, a recently stored message, which now corresponds to a clique with a **unique** tag in the network, can overwrite parts of an older message they share by changing the tag of those shared connections.

Another, more visual, formulation using colored graph theory is that tags can be seen as different overlays or colors of the same network, each defining a sub-graph containing edges of only one color. By focusing on one color, it's easy to decode the clique without ambiguity, which is not possible with the original clique network. Thus, those colored layers can be separated easily: newer messages can be retrieved flawlessly, while older messages can still be retrieved without ambiguity, as illustrated by Figure 3.

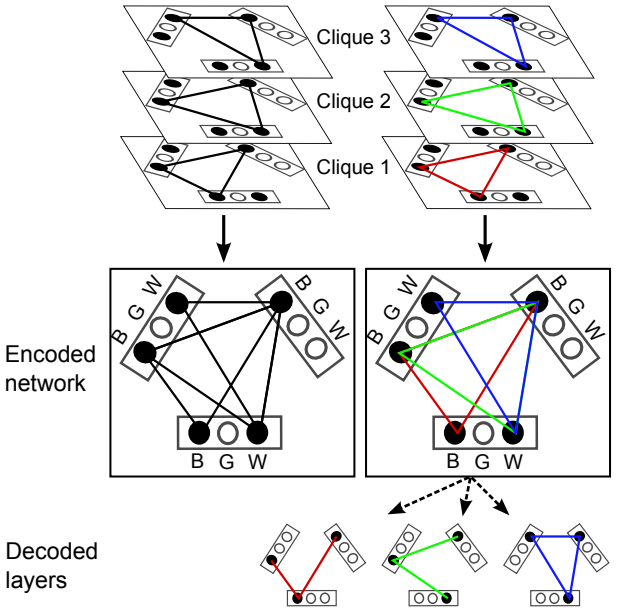


Figure 3. Comparison of the clique network (left) and tagged network (right), with tags represented as colored layers.

More formally, let us consider that messages to store are assigned a tag k from 1 to g. The storing process can be defined as constructing the adjacency matrix A of the colored graph, but instead of assigning 0 or 1 to assert an edge existence, we assign the latest, highest tag k attached to each edge:

$$A_{(ij)(i'j')} = \{max(k) \big| \exists m \in \mathcal{M}, \\ tag(m) = k \wedge m_i = j \wedge m_{i'} = j'\}, \\ \text{for } 1 \leq i, i' \leq \chi \text{ and } 1 \leq j, j' \leq \ell. \quad (2)$$

Note that k is to be defined by an assignment function, which purpose is to generate a tag for each message. We will later in this section discuss about several possible strategies.

We thus obtain an adjacency matrix where entries are valued from 0 (edge nonexistence) to g (existence + tag membership).

In order to benefit from this added material, the iterative retrieval process is adapted, as shown in Figure 4.

1) Decode just like before: propagate using Sum-Of-Sum rule and filter using Global Winners-Take-All rule.
2) Disambiguation post-processing step:
   For each message:
   a) Find major tag (= compute mode) among edges.
   b) Delete every edges possessing a different tag than the major.
   c) Delete isolated units.

Figure 4. Global-Vote-Local-Elimination retrieval rule

This results in a collaborative decision between units, which will favor likely tagged edges, and remove units that don't share at least one edge of the correct, major tag.

This combination of a local elimination based on a global, cooperative decision is the most successful strategy, which we call the *Global-Vote-Local-Elimination rule*. We also tried several other variants like a local vote (compute mode per each node) and global elimination (filter out all nodes which local major tag isn't the global major tag) but they all produced significantly lower performance. The voting strategy to find the most likely tag is probably optimal, since we simulated a tagged network with tag guiding (the tag for each clique is known, hence there's no uncertainty), and it provided no difference in performances.

Of course, assigning a tag per clique is optimal, since the tag is then unique. However, $\log_2(g)$ more resources than the clique network are needed to store the tags information. Hence, it's possible to tradeoff with the number of tags g compared to the total number M of messages in the set $\mathcal{M}$:

- $g = 1$ will output the same result as the non-tagged clique network, since all edges will have the same tag, the tags disambiguation step will just have no effect.
- $1 < g < M$ defines a limited set of tags to use among all cliques. This produces a trade-off between network's capacity and the amount of resources required to represent the tags. In practice, since there is a limited set of tags available, they will be recycled among the cliques, thus rendering tags non-unique and producing more ambiguities. Any surjective function can be used to map the tags onto the cliques. In practice, it seems reasonable to use a uniform distribution, which will distribute randomly the tags almost uniformly among messages, and also allow for *online learning* (learning new messages over time).
- $g = M$ will assign one unique tag per clique. In this case, we get as many tags as there are cliques. Performance is then optimal, but more resources are consumed. In such case we consider that the set of messages to store are ordered from lowest to highest tag, such that a message with a high tag number is said to have been stored recently and a message with a low tag is said to be old.

IV. Analysis

A. Density

Since the network still relies on cliques and edges existence to store and retrieve information, the theoretical density d – defined as the ratio of used edges to that of possible ones– is just the same as in the classical clique network [8]:

$$d = 1 - \left(1 - \frac{c(c-1)}{\chi(\chi-1)\ell^2}\right)^M . \tag{3}$$

B. Efficiency

The network is split into χ parts with ℓ units in each. Thus, the network possess $n = \chi\,\ell$ total units and $\frac{\chi(\chi-1)\ell^2}{2}$ total possible edges. Furthermore, edges aren't binary anymore, but store their tag, thus an edge can now store a value between 0 and g, and therefore the tagged network representation amounts to a binary resource Q of:

$$Q = \frac{\chi(\chi-1)\ell^2}{2}\,\log_2(g+1)\,. \tag{4}$$

The entropy per message b and total entropy B, or amount of binary information B learned by the network after storing all message M, does not change from the classical model [8]:

$$b = \log_2\binom{\chi}{c} + c\,\log_2(\ell)\,. \tag{5}$$

$$B = bM = M\left(\log_2\binom{\chi}{c} + c\,\log_2(\ell)\right). \tag{6}$$

We can then easily derive the network's efficiency η, that is the efficient usage of available network's resources:

$$\eta = \frac{B}{Q} = \frac{2M\left(\log_2\binom{\chi}{c} + c\,\log_2(\ell)\right)}{\chi(\chi-1)\,\ell^2\cdot\log_2(g+1)}\,. \tag{7}$$

which leads to the network's *efficiency-1 diversity*, an upper bound of the optimal number of M messages to store to maximize the network's efficiency:

$$M_{max} = \frac{Q}{b} = \frac{\chi(\chi-1)\ell^2\cdot\log_2(g+1)}{2\left(\log_2\binom{\chi}{c} + c\,\log_2(\ell)\right)}\,. \tag{8}$$

 ISBN: 978-1-61208-390-2

C. Error rate

Since we are doing an associative task, that is we are trying to recover a full corrected message from a query, a *retrieval error* is defined as the network converging to a different, *spurious clique* than the correct clique from which the (partial, noisy or complete) input query was generated from.

Let us now suppose that we set $g = M$. As described in the previous section, at the disambiguation step of the decoding process, only nodes without any edge of the major tag will be filtered out. This implies that even if an old edge from an old clique can get its tag overwritten by a new clique, the two units, which the edge is linked to, are still retrievable without a hitch as long as they each possess at least one other edge with the proper tag. This means that for a clique to be irretrievable anymore, it has to lose at least one unit, and to lose one unit is for this unit to be shared by so many other, new cliques that *all edges of this unit got overwritten*.

This kind of error, very specific to the tagged network, is what we call the *lost unit error* P_{lost}, and is the most significant factor contributing to retrieval error. It is also quite interesting for the fact that it's only defined by the learning process (new cliques overwriting tags of old cliques' edges), without any influence by the decoding dynamics.

Since the network store cliques of size (number of edges) $\frac{c(c-1)}{2}$, and if we consider that cliques' edges are generated randomly uniformly among all possible network's edges, then the probability to overwrite one edge of an old unit when storing a new clique is $\frac{2}{\chi(\chi-1)\ell^2}$, and we can define P_{lost} as follows:

$$P_{lost} = \left(1 - \left(1 - \frac{2}{\chi(\chi-1)\ell^2}\right)^{(M-1)\frac{c(c-1)}{2}}\right)^c . \quad (9)$$

This formula can be understood as the probability that an edge gets overwritten by any edge from every learnt messages, and then this probability is powered to c because for a unit to be lost, every one of its edges belonging to the original clique must be overwritten. This formula is but an approximation of the lost unit error because of our assumption that messages are *i.i.d.*, which is of course not the case.

Beside the lost unit error, other factors may contribute to a wrong retrieval, such as when the vote to compute the major tag leads to a wrong tag (*tag vote error*), or when the propagation/filtering steps lead to a wrong clique (leading to a spurious clique, just like with the classical clique network) before the tag disambiguation step (*decoding error*). Yet, after simulating the tagged network's dynamics with erasures, we have found that the lost unit error is prevalent, and is a very good approximation of the overall error, as shown in Figure 5. However, this is only true when $g = M$, if we use only a limited definite number of tags lower than M, other types of error will have increasingly more effect.

V. Simulations

We analyze the network's performance with an erasure scenario, which is the substitution of one or several nonzero

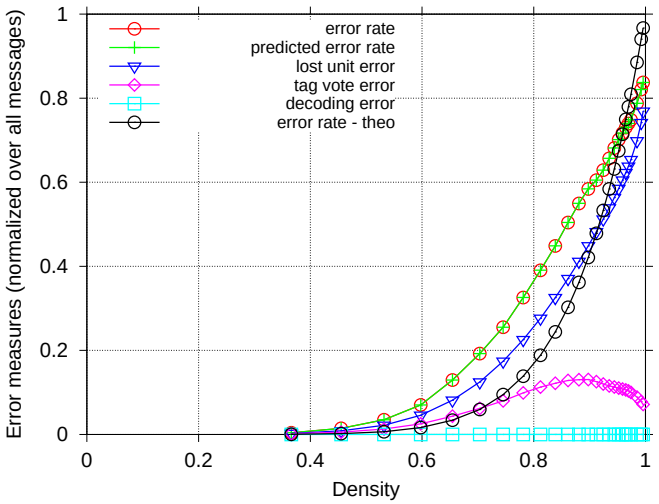


Figure 5. Impact of various types of errors on message retrieval in a sparse tagged network. The errors are computed as the ratio of total messages suffering from this particular error type over all learnt messages. Predicted error rate is the sum of all error types to check that they are good predictors of the real error rate. As can be seen, the lost clique error is a very close predictor of the real error rate, with a small difference due to the impact of the tag vote error.

symbols in a message m by 0. The simulations were done by learning uniformly random messages and then each point was generated by sampling a subset of the learnt messages and erase half of the nonzero symbols. To avoid random fluctuations, each point has been averaged over 10k trials (200 messages per 50 different networks).

To study the influence of tags on the error rate, we simulated a sparse tagged network with various numbers of tags, against a classical sparse clique network with similar parameters as described by Aliabadi et al. [8]. As can be seen in Figure 6, tags greatly impact on network's performance, significantly lowering the error rate even with a small finite set of tags such as 5, but the maximal gain is of course obtained by using M tags (one unique tag per message).

Empirically, we found that to maximize the tagged network's performance, some key parameters need to be set, in particular: the network must be sparse ($\chi > c$) ; there must be more than one unit per part ($\ell > 1$) ; and the γ memory effect [8] should be set to 1.

Thus, tags significantly enhance the retrieval process, but to be fair, we have to consider the added resources we use in the network to account for those tags. Therefore, we have done a similar simulation in Figure 7, but we here compared the clique network's efficiency with the tagged network's, and to set a comparable frame we plot the efficiencies with respect to the error rate since the tagged network can run at far higher density regimes than the clique network.

This shows that tags are a less efficient mean to extend a network's diversity than by resizing the network. However, these two means are not exclusive, and thus can be used concurrently to extend a network: first by size up to a limit, and then tags can be used to extend further. This can be a

 ISBN: 978-1-61208-390-2

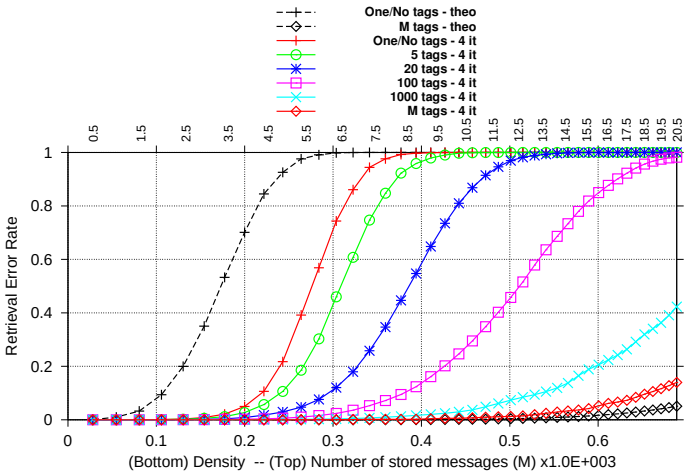


Figure 6. Error rate evolution with respect to the number of maximum tags allowed to be assigned to learned cliques. "One tag" curve corresponds to the sparse clique network [8] and serves as a reference, while "M tags" is when a unique tag is assigned per message. Network's parameters are $\chi = 16$, $c = 8$, $\ell = 64$, erasure rate $\alpha = 0.5$ (half of the c units are erased from input query) and 4 decoding iterations. The plot has two axes: the main one at the bottom defines the network's density (how much the network is full of messages), which eases comparison between different figures because the density isn't influenced by network's size, while the second axis at the top is the number of learnt messages this density corresponds to.

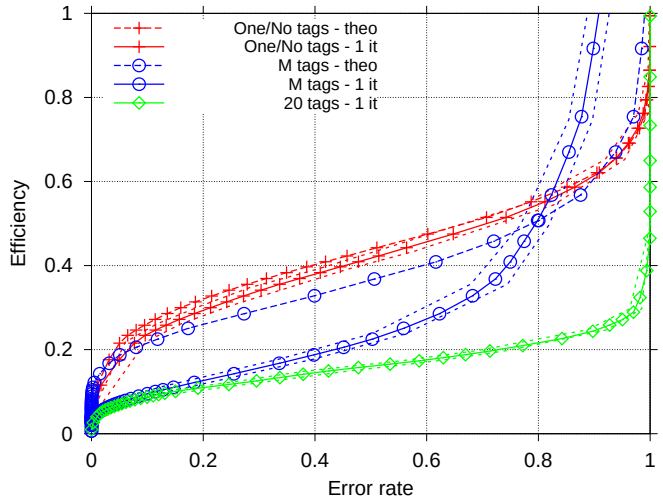


Figure 7. Efficiency with respect to the error rate for the sparse tagged network (green) compared to the clique network (red). Standard deviations are shown in dotted lines and theoretical error rates in dashed lines. Network's parameters are the same as in Figure 6 except that we here use only 1 decoding iteration. Note that the curves remain identical whatever the network's size is.

very interesting alternative for devices with a finite, static set of units, such as VLSI (Very-Large-Scale Integration) [19]–[21], ASIC (Application-Specific Integrated Circuit) [22] or FPGA (Field-Programmable Gate Array) [23] based neuromorphic and neuromemristive [24] hardware devices, where it is certainly easier to add a tag counter than to resolder the board in order to resize the network.

These results are reproducible via the complete source-code in MatLab/Octave, which is freely available online [25].

VI. Discussions

The biological mesoscopic mechanisms of learning and memory storing are extremely complex and are still a mystery. The mechanisms of forgetting are even further from grasp. This simple extension provides an elegant way to implicitly implement a forgetting mechanism with variable effect (stronger when the number of tags is low), and as such this provides a continuum to transition from a long-term memory model (clique network), where there's no overwriting nor "forgetting" of old memories, to a palimpsestic working memory [17], [26], [27]. However, even long term memory can benefit from forgetting, as this fundamental regulating mechanism seems to be tightly coupled with the retention of memories in order to mitigate the overfitting (lack of generalization) phenomenon [28]–[31]. An interesting side effect is that refreshing tags on access (i.e., when a clique is accessed, the computed major tag is assigned to each of its edges, thus "refreshing" the clique) could account for the spacing effect in learning [32]–[34], and future work in this direction may yield interesting results.

On a biological side, there is currently no observation of such an implementation of tags, and we don't argue that tags are physically embodied as-is in a biological brain. However, the tags model a concept that is far more general: affinity between synapses. Biological mechanisms behind such affinities are still merely assumptions, yet they are not implausible: synaptic discrete states [16], resonance, alike morphologies, synapse's conductance rate using variable myelination [35], [36], cascading biochemical signature [37], or a sensory modality cue. This is not as far-fetched as it sounds, as it is currently thought, according to the synaptic tagging model [38]–[42], that the very process of memory creation uses some kind of chemical tagging to convert recent, short-term and weak, memories into long-term, long-lasting and resilient, memories.

The following is merely a hypothesis, but if we suppose that the brain is ruled by stochastic processes, then if a set of synapses get created at the same moment – which may certainly be the case if synaptogenesis can be triggered in a synchronized way by glial cells just like they can trigger the synchronization of synaptic communications over wide areas [36], [43] – these synapses may get the same identical set of specific parameters (since they were created at the same moment of the stochastic process ruling the parameters of synaptic parameters), whether those parameters are a similar chemical signature, a similar myelination profile (which is very heterogeneous, probably unique, for each synapse over the brain), a similar activation threshold, a similar set of neurotransmitters, or just a similar morphology. Sharing a similar set of attributes may allow these synapses to mutually sustain, to resonate, when one or more of these synapses are activated at the same moment later in time, as some kind of reminder that they also were created together. This could be seen as some sort of evolutionary collaboration: these synapses were created at the same moment, and thus probably embody

 ISBN: 978-1-61208-390-2

some sort of co-occurrence in information, and an affinity may perfectly encode that.

Let's now discuss about how a tagged network could technically and efficiently be implemented in neuromorphic hardwares. We mentioned in the previous section a few neuromorphic technologies that could benefit from tags, but a specific type of component could be the most efficient way to achieve a very low-energy neuromorphic device based on tags: memristors, and in particular compound memristors [24], could be a great fit for tags since they can store finite precision integers, while retaining the very interesting feature of any memristor, that is to use almost no current to maintain their state. The compound memristors could thus be used to store edges tags at a very low energy cost, which is sufficient and enough to define the storage of a whole neural network based on tags.

VII. Conclusion and future work

We have presented a new generic method to extend an associative memory network's diversity by tags, which provides an alternative to the network's size versus messages diversity trade-off. We based this method on an efficient associative memory model called clique neural network, and we provided the algorithmics underpinning this extension, which we called the *tagged neural network*. We then analyzed the network's dynamics and simulated a retrieval scenario with partially erased queries in order to study the impact on performance and efficiency of this extension, which demonstrate that tags can be used as a viable alternative, although a bit less efficient, to extend an associative memory neural network's capacity when the network's size is constrained.

Future work on this approach should focus on the analysis in a noisy scenario, where tags would not be reliable indicators anymore. In this scenario, it may be advisable to adapt the retrieval process to account for this uncertainty of the tag indicator. Another interesting avenue is the fact that the biggest source of decoding error in this network resides in the tag overwriting of old cliques by new cliques, resulting in the lost units error we described. This source of error may potentially be reduced by adapting, interestingly, the learning process, and not the decoding process, since losing units happens at the learning stage, without any influence of the decoding stage. To be more explicit: the biggest source of error is structurally encoded in the network at the learning stage, thus, optimization effort should focus on the learning process.

Also, tags are flexible indicators, whose underlying representation is totally dependent on the designer's conception. This flexibility of representation can be used for various purposes and applications beyond neuromorphic hardware, for example, by using tags as semantic cues: a tag can be seen as a label representing an identity/class of the clique pattern. Hence, a tagged network may not only increase storage diversity but also be seen as a clear identification system for specific kinds of patterns. Thus, the same tag could be used to regroup patterns that are semantically similar, or which originate from the same sensory modality (e.g., using the same tag to regroup all patterns originating from vision, another tag for audio, another one for taste, etc.). If efficient enough, this semantic use of tags could be applied successfully to a wide array of applications where we need to semantically disambiguate, such as objects class recognition in a scene.

Acknowledgment

This work was partially funded as part of the NEUCOD project by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement n° 290901.

References

[1] C. Frith and R. Dolan, "The role of the prefrontal cortex in higher cognitive functions," Cognitive brain research, vol. 5, no. 1, 1996, pp. 175–181.

[2] J. Von Neumann, The computer and the brain. Yale University Press, 1974.

[3] W. Aspray, John von Neumann and the origins of modern computing. Mit Press Cambridge, MA, 1990, vol. 191.

[4] L. C. Aiello and P. Wheeler, "The expensive-tissue hypothesis: the brain and the digestive system in human and primate evolution," Current anthropology, 1995, pp. 199–221.

[5] C. W. Kuzawa et al., "Metabolic costs and evolutionary implications of human brain development," Proceedings of the National Academy of Sciences, Aug. 2014, p. 201323099.

[6] A. Knoblauch, G. Palm, and F. T. Sommer, "Memory capacities for synaptic and structural plasticity," Neural Computation, vol. 22, no. 2, 2010, pp. 289–341.

[7] V. Gripon and C. Berrou, "Sparse neural networks with large learning diversity," Neural Networks, IEEE Transactions on, vol. 22, no. 7, 2011, pp. 1087–1096.

[8] B. K. Aliabadi, C. Berrou, V. Gripon, and J. Xiaoran, "Storing sparse messages in networks of neural cliques," IEEE transactions on neural networks and learning systems, vol. 25, no. 5, 2014, pp. 980–989.

[9] A. Aboudib, V. Gripon, and X. Jiang, "A study of retrieval algorithms of sparse messages in networks of neural cliques," in COGNITIVE 2014, The Sixth International Conference on Advanced Cognitive Technologies and Applications, 2014, pp. 140–146.

[10] S. Kurt et al., "Auditory cortical contrast enhancing by global winner-take-all inhibitory interactions," PLoS ONE, vol. 3, no. 3, Mar. 2008, p. e1735.

[11] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," Proceedings of the national academy of sciences, vol. 79, no. 8, 1982, pp. 2554–2558.

[12] D. J. Willshaw, O. P. Buneman, and H. C. Longuet-Higgins, "Non-holographic associative memory," Nature, vol. 222(5197), June 1969, pp. 960–962.

[13] T. Kojima, H. Nonaka, and T. Da-Te, "Capacity of the associative memory using the boltzmann machine learning," in Neural Networks, 1995. Proceedings., IEEE International Conference on, vol. 5, Nov 1995, pp. 2572–2577 vol.5.

[14] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural computation, vol. 9, no. 8, 1997, pp. 1735–1780.

[15] V. Gripon, "Networks of neural cliques," Ph.D. dissertation, Télécom-Bretagne, Jul. 2011.

[16] J. M. Montgomery and D. V. Madison, "Discrete synaptic states define a major mechanism of synapse plasticity," Trends in Neurosciences, vol. 27, no. 12, 2004, pp. 744–750.

[17] J. Sacramento and A. Wichert, "Binary willshaw learning yields high synaptic capacity for long-term familiarity memory," Biological cybernetics, vol. 106, no. 2, 2012, pp. 123–133.

[18] A. M. Dubreuil, Y. Amit, and N. Brunel, "Memory capacity of networks with stochastic binary synapses," PLoS computational biology, vol. 10, no. 8, 2014.

[19] H. University. Brainscales - neuromorphic processors. [Online]. Available: http://www.artificialbrains.com/brainscales [retrieved: 03, 2015]

[20] D. Brüderle et al., "A comprehensive workflow for general-purpose neural modeling with highly configurable neuromorphic hardware systems," Biological cybernetics, vol. 104, no. 4-5, 2011, pp. 263–296.

 ISBN: 978-1-61208-390-2

[21] G. Indiveri, "Neuromorphic bistable vlsi synapses with spike-timing-dependent plasticity," in NIPS, 2002, pp. 1091–1098.
[22] L. S. Smith, "Neuromorphic systems: past, present and future," in Brain Inspired Cognitive Systems 2008. Springer, 2010, pp. 167–182.
[23] A. Cassidy, A. G. Andreou, and J. Georgiou, "Design of a one million neuron single fpga neuromorphic system for real-time multimodal scene analysis," in Information Sciences and Systems (CISS), 2011 45th Annual Conference on. IEEE, 2011, pp. 1–6.
[24] J. Bill and R. Legenstein, "A compound memristive synapse model for statistical learning through stdp in spiking neural networks," Frontiers in Neuroscience, vol. 8, no. 412, 2014, pp. 1–18. [Online]. Available: http://www.frontiersin.org/neuromorphic_engineering/10.3389/fnins.2014.00412/abstract
[25] S. Larroque. Clique network (gbnn) implementation in octave/matlab. (doi: http://dx.doi.org/10.5281/zenodo.15788). [Online]. Available: https://github.com/lrq3000/gbnn-matlab [retrieved: 03, 2015]
[26] G. Parisi, "A memory which forgets," Journal of Physics A: Mathematical and General, vol. 19, no. 10, 1986, p. L617.
[27] D. J. Amit and S. Fusi, "Learning in neural networks with material synapses," Neural Computation, vol. 6, no. 5, 1994, pp. 957–982.
[28] A. M. Jasnow, P. K. Cullen, and D. C. Riccio, "Remembering another aspect of forgetting," Frontiers in psychology, vol. 3, 2012, p. 175.
[29] C. O'Donnell and T. J. Sejnowski, "Selective memory generalization by spatial patterning of protein synthesis," Neuron, vol. 82, no. 2, 2014, pp. 398–412.
[30] A. Dovgopoly and E. Mercado III, "A connectionist model of category learning by individuals with high-functioning autism spectrum disorder," Cognitive, Affective, & Behavioral Neuroscience, vol. 13, no. 2, 2013, pp. 371–389.
[31] R. Spencer, "Neurophysiological basis of sleep's function on memory and cognition," ISRN Physiology, vol. 2013, 2013, p. 17.
[32] H. Ebbinghaus, "Memory: A contribution to experimental psychology," Annals of Neurosciences, vol. 20, no. 4, 10 2013, pp. 155–156, [Über das gedächtnis: untersuchungen zur experimentellen psychologie. Duncker & Humblot, 1885].
[33] F. N. Dempster, "The spacing effect: A case study in the failure to apply the results of psychological research." American Psychologist, vol. 43, no. 8, 1988, p. 627.
[34] R. A. Bjork, "Assessing our own competence: Heuristics and illusions." 1999.
[35] G. S. Tomassy et al., "Distinct profiles of myelin distribution along single axons of pyramidal neurons in the neocortex," Science, vol. 344, no. 6181, 2014, pp. 319–324.
[36] N. Levine-Small, K. Mueller, R. Guebeli, B. Chow, W. Weber, and U. Egert, "Selective stimulation of astrocytes modulates activity states in neuronal networks," 2014.
[37] S. Fusi, P. J. Drew, and L. Abbott, "Cascade models of synaptically stored memories," Neuron, vol. 4, no. 45, 2005, p. 45.
[38] U. Frey and R. G. Morris, "Synaptic tagging and long-term potentiation," Nature, vol. 385, no. 6616, 1997, pp. 533–536.
[39] C. Clopath, L. Ziegler, E. Vasilaki, L. Büsing, and W. Gerstner, "Tag-trigger-consolidation: A model of early and late long-term-potentiation and depression," PLoS Computational Biology, vol. 4, no. 12, 12 2008, p. e1000248.
[40] C. Clopath, "Synaptic consolidation: an approach to long-term learning," Cognitive neurodynamics, vol. 6, no. 3, 2012, pp. 251–257.
[41] S. Sajikumar, S. Navakkode, and J. U. Frey, "Identification of compartment-and process-specific molecules required for "synaptic tagging" during long-term potentiation and long-term depression in hippocampal ca1," The Journal of neuroscience, vol. 27, no. 19, 2007, pp. 5068–5080.
[42] S. Frey and J. U. Frey, "'synaptic tagging'and 'cross-tagging'and related associative reinforcement processes of functional plasticity as the cellular basis for memory formation," Progress in brain research, vol. 169, 2008, pp. 117–143.
[43] N. Levine-Small et al., "Astrocytes drive neural network synchrony," in MEA Meeting, 2012, p. 30.

 ISBN: 978-1-61208-390-2

Two Minds and Emotion

Muneo Kitajima

Nagaoka University of Technology
Kamitomioka, Nagaoka, Niigata JAPAN
Email: mkitajima@kjs.nagaokaut.ac.jp

Makoto Toyota

T-Method
Sapporo, Hokkaido JAPAN
Email: pubmtoyota@me.com

***Abstract*—Human behavior can be viewed as the integration of the outputs of Systems 1, i.e., unconscious automatic processes, and System 2, i.e., conscious deliberate processes. System 1 activates a sequence of automatic actions. System 2 monitors System 1's performance according to the plan it has created, and it activates future possible courses of actions as well. At the same time when these forward processes are working, System 1 and System 2 deal with the outcome of the forward processes by estimating the results of System 1's and System 2's performance, i.e., good or bad, and generating emotions depending on the degree of goodness or badness of the estimation. Emotions are generated through the dynamics of the parallel processing of System 1 and System 2, which is called O-PDP, Organic Parallel Distributed Processing. This paper discusses how emotion generation process is integrated with the Kahneman's System 1 and System 2 model of human decision-making.**

Keywords–Two Minds; Organic-PDP; Two Minds; Consciousness; Unconsciousness; Emotion.

I. INTRODUCTION

Two Minds is the basis of behavioral economics founded by Kahneman [1][2]. It considers that Two Minds govern human decision-making: a human being's behavior is the outcome of two different systems including an experiential processing system, System 1, and a rational processing system, System 2. Figure 1 illustrates the workings of the two systems. In short, System 1 is a fast feed-forward control process driven by the cerebellum and oriented toward immediate action. Experiential processing is experienced passively, outside of conscious awareness, e.g., one is seized by one's emotion. In contrast, System 2 is a slow feedback control process driven by the cerebrum and oriented toward future action. It is experienced actively and consciously, e.g., one intentionally follows the rules of inductive and deductive reasoning. However, traditional cognitive modeling has not treated human behavior as the result of intense interaction between System 1 and System 2. Indeed, traditional cognitive modeling has been primarily aiming at implementing and testing theories explaining behavior driven by rational, multi-step cognition. ACT-R [3][4] and Soar [5] are the most successful cognitive architectures in this direction. However, there are many human behaviors that seem to be driven by aspects of behavior that are not the same as rational cognition: emotional, intuitive, and other non-rational behavior. These aspects have not been addressed in these cognitive architectures adequately.

Recently, studies aiming at incorporating Two Minds into traditional cognitive modeling have emerged. For example, Kennedy and Bugajuska [6] presented computational cognitive models that take into account the interactions between conscious processes and unconscious processes. Their models implemented different strategies in the ACT-R cognitive architecture [7][4] to allow a human to consciously inhibit an undesirable fast response. ACT-R is a symbolic and subsymbolic, production-based cognitive architecture [3]. The internal modules of ACT-R represent relatively specific cognitive functions (and regions of the brain) including declarative and procedural memory (long-term memory), auditory and visual perception, and vocalization and motor functions.

This paper suggests another approach to incorporating Two Minds in cognitive modeling by stemming on the cognitive architecture, Model Human Processor with Realtime Constraints (MHP/RT), the authors have developed [8][9]. MHP/RT considers that human behavior in the real world occurs in such a way that the agent's next behavior is determined by a combination of the situation of the environment and that of the agent itself. Therefore, in order to simulate the processes of human behavior, it is necessary to explicitly include the real-time constraints that affect the synchronization of behavior selection. Neither Soar nor ACT-R was designed as an architecture model for simulating an agent's behavioral processes that evolve in synchrony with the environment where real-time constraints are the critical factors to organize the agent's behavior. The predictions made by these architecture models are derived essentially by linear algorithms that calculate the best paths for a sequence of behavioral selections [8].

As described in [9], this feature of MHP/RT should be contrasted with the goal-oriented cognitive architectures such as ACT-R [3][4] in which the conscious processes are considered as the processes to control people's behavior and the unconscious processes are considered subordinate to the conscious or intentional processes. What ACT-R tries to do is to show how System 2 can be implemented on top of System 1. The procedural memory system is very similar to System 1 (fast, learning based on rewards/experience, intuitive), and then ACT-R models tend to consist of a set of production rules that (when run on this System 1 module and in combination with symbolic working memory buffers and a long-term memory system) give rise to the slower, deliberative planning behaviors seen in System 2. This is a very different approach to MHP/RT. However, ACT-R models are totally adequate for simulating stable human activities with weak time constraint in which deliberate decision making would work effectively, but might be hard for the situations with strong time constraint where the environmental condition changes chaotically and deliberate decision making implemented on System 2 might not work as effective.

 ISBN: 978-1-61208-390-2

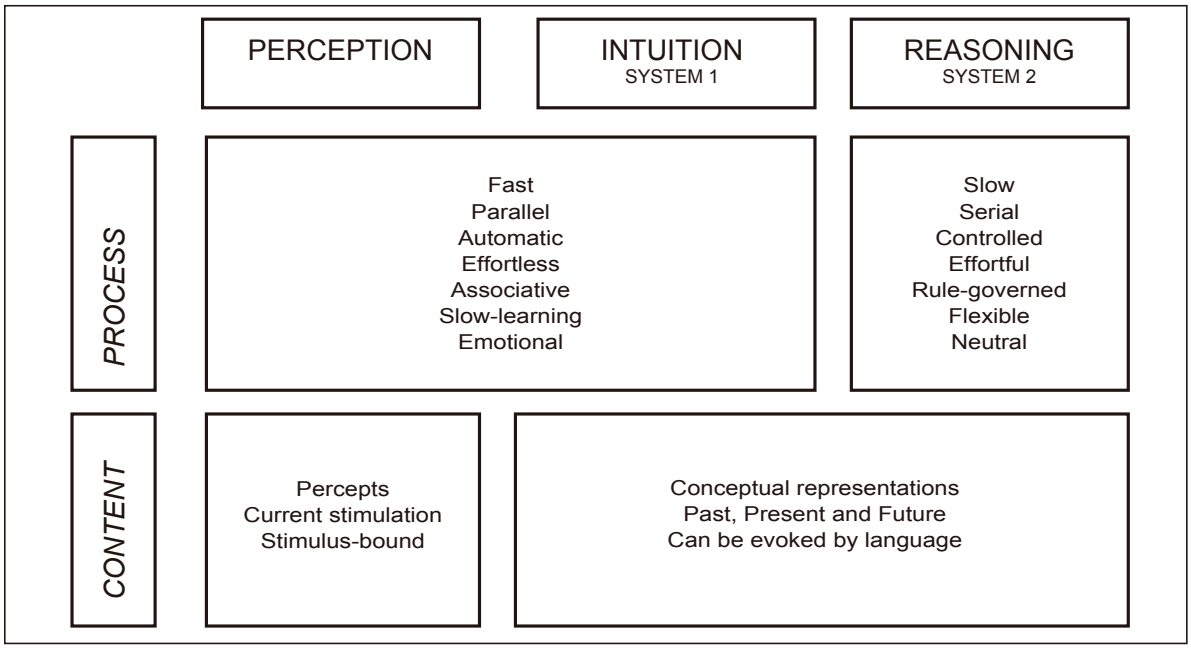


Figure 1. Process and Content in Two Cognitive Systems (adapted from [1]).

In summary, traditional cognitive modeling, including ACT-R or Soar based models, has not considered how System 1 and System 2 develop along the time dimension in synchronous with the ever-changing environment. Especially, the intense interaction between emotion in System 1 and consciousness in System 2 has not been considered appropriately due to the lack of proper treatment of the time dimension. This paper discusses how *emotion generation processes* are integrated with Two Minds on the basis of MHP/RT [8][9] and its superordinate model, the Nonlinear Dynamic Human Behavior Model with Real-Time Constraints (NDHB-Model/RT), that the authors proposed at the past Cognitive Science conferences [10][11][12].

This paper is organized as follows. In Section II, this paper starts by describing NDHB-Model/RT, followed by an explanation of dynamics of consciousness–emotion interaction based on NDHB-Model/RT in Section III. Section IV describes relationship between Two Minds and emotions, and Section V concludes the paper.

II. NDHB-MODEL/RT

The relationship between Two Minds and emotion is best understood by considering how behavior is generated in the time dimension. This section introduces some ideas to make it possible. First, it introduces two basic ideas, Organic Self-Consistent Field Theory (O-SCFT), and Organic Parallel Distributed Processing (O-PDP). Then it derives Nonlinear Dynamic Human Behavior Model with Realtime Constraints (NDHB-Model/RT), which is a framework for considering the behavior of human beings in the universe as defined by O-SCFT and O-PDP. Then, it introduces a cognitive architecture, Model Human Processor with Realtime Constraints (MHP/RT), that is capable of simulating human being's daily decision making and action selection under NDHB-Model/RT.

A. O-SCFT: Organic Self Consistent Field Theory

1) Self-Consistent Field Theory in Physics: In physics, self-consistent field theory (SCFT) studies the behavior of large and complex stochastic models by studying a simpler model. Such models consider a large number of small interacting individual components which interact with each other. The effect of all the other individuals on any given individual is approximated by a single averaged effect, thus reducing a many-body problem to a one-body problem. In field theory, the Hamiltonian may be expanded in terms of the magnitude of fluctuations around the mean of the field. In this context, SCFT can be viewed as the "zeroth-order" expansion of the Hamiltonian in fluctuations. Physically, this means an SCFT system has no fluctuations, but this coincides with the idea that one is replacing all interactions with a "self-consistent field." Quite often, in the formalism of fluctuations, SCFT provides a convenient starting-point to studying first or second order fluctuations.

2) "Organic" Self-Consistent Field Theory: We applied SCFT in physics to organic systems. Organic systems are those comprised of human beings as their components. Any organic system can be represented as a model that considers a large number of interacting individual human beings which interact with each other. In addition, individual *organic* human beings interact with *inorganic* physical environment as well, which is modeled by SCFT. We prefixed the word "Organic" to SCFT in order to explicitly indicate that the application domain of SCFT is extended to organic systems. We consider that the behavior of human beings in the universe is quasi-stable, which means that it is not stable but develop or evolve triggered by some fluctuations, a feature of dissipative system – a fluctuation of the system caused by an environmental change would trigger creation of a new order or catastrophe [13].

3) Human beings considered in O-SCFT: At the zeroth-order approximation implied by O-SCFT, each human being interacts with the integrated environment consisting of inorganic components and organic components. Each human being

is considered as *autonomous system*, and interaction is best represented by *information flow* from the view point of human being. O-SCFT is decomposed into three nonlinear constructs, Maximum Satisfaction Architecture (MSA), Brain Information Hydrodynamics (BIH), and Structured Meme Theory (SMT) that correspond to human being, inorganic SCFT components, and O-SCFT components, respectively.

a) Maximum Satisfaction Architecture (MSA): MSA is about realization of the purpose of living, i.e., libido – it maximizes efforts on the autonomous system. It deals with how autonomous systems achieve goals under constraints defined by BIH and SMT [10].

b) Brain Information Hydrodynamics (BIH): Constrains from the environment shape how the information flow develops along the time dimension. This is reflected in the brain as BIH. It deals with information flow in the brain and its characteristics in the time dimension [11].

c) Structured Meme Theory (SMT): SMT concerns the relational structure that links human beings and the environment, and thereby deals with effective information and the range of propagation [12].

B. O-PDP: Organic Parallel Distributed Processing

We are interested in not only how individual human being's brain processes information, originated either from external or internal environment, but also how it develops chronologically from his/her birth. We challenge this problem under the concepts of MSA, BIH, and SMT. As we focus on information flow in the brain, we considered that Parallel Distributed Processing (PDP) is the fundamental mechanism for developing brain architecture [14]. Since PDP is considered under O-SCFT, we prefixed "O (organic)" to PDP, and call this approach O-PDP.

O-PDP develops cross-networks of neurons in the brain as it accumulates experience of interactions in the environment. The neural network development process is *circular*, which means that any experience at a particular moment should reflect somehow the experience of the past interactions that have been recorded in the shape of current neural networks. In this way, a PDP system is organized evolutionally, and realized as a neural network system, including the brain, the spinal nerves, and the peripheral nerves to construct an O-PDP system. This mechanistic statement might be expressed simply and casually with a less rigorous but easier to understand way as follows: the brain processes information thanks to a dynamically reconfigurable network of neurons to which one must add the spinal chord and peripheral nerves.

C. NDHB-Model/RT

On the basis of O-SCFT and O-PDP, we have developed NDHB-Model/RT as an architecture model that consists of a behavioral processing system and a memory processing system that interact with each other as autonomous systems. The interactions are cyclic, and memory develops and evolves as time goes by. NDHB-Model/RT represents *consciousness* as one-dimensional linear operations, i.e., language, corresponding to System 2 of Two Minds, and *unconsciousness* including emotion as a hydrodynamic flow of information in multi-dimensional parallel operations in the neural networks, corresponding to System 1 of Two Minds. NDHB-Model/RT has autonomous memory systems that mediate between consciousness and unconsciousness to display the dynamic interactions between them.

NDHB-Model/RT suggests that the brain consists of the following three non-linearly connected layers. Behavioral decisions and action selections are made by integrating the results of operations of these three layers:

- C-layer: Conscious state layer, i.e., System 2 of Two Minds
- A^2BC-layer: Autonomous-Automatic Behavior Control layer, i.e., System 1 of Two Minds
- B-layer: Bodily state layer

B-layer prioritizes the 17 behavioral goals, i.e., happiness types defined by [15], such as "target happiness for an achiever", "cooperative happiness for a helper", "rhythmic happiness for a dancer", and so on. The other two layers interact with each other in order to derive the next behavior that should satisfy the highest prioritized goal. In normal situations in our daily life, temporal changes in the environment impose the strongest constraint on the decision of the next behavior, and thus A^2BC-layer plays a more dominant role than C-layer in organizing behavior. To put it simply, in our daily life we act more by reflex than by reasoning.

The next behavior is determined by extracting objects from the ever-changing environment and attaching values to them according to the degree of the strength of the resonance with what is stored in the autonomic memory system. This is followed by deliberate judgement by using the knowledge associated with the highly valued objects. The former is controlled by the processes in A^2BC-layer, System 1; the latter, by the processes in C-layer, System 2.

D. MHP/RT

NDHB-Model/RT can be simulated by the architecture model, Model Human Processor with Real time Constraints (MHP/RT) [8][9][16]. MHP/RT simulates in situ human behavior by switching among four processing modes, conscious/unconscious activities for the future events, and conscious/unconscious activities for the past events. MHP/RT focuses on synchronization between System 1 and System 2 in the information flow under O-PDP. More specifically, MHP/RT deals with one aspect of working of NDHB-Model/RT, which is synchronization between conscious system and unconscious system in the ever-changing environment where human beings make decisions and action selections to behave properly.

Figure 2 depicts the outline of MHP/RT. It is a *real* brain model comprising of System 1's unconscious processes and System 2's conscious processes at the same level. There are two distinctive information flows: System 1 and System 2 receive input from the Perceptual Information Processing System in one way, and from the Memory Processing System in another way. System 1 and System 2 work autonomously and synchronously without any superordinate-subordinate hierarchical relationships but interact with each other when necessary. In Figure 2, solid lines and dotted lines indicate the path associated with System 1 and the one associated with System 2, respectively. These two flows are synchronized before carrying out some behavior.

 ISBN: 978-1-61208-390-2

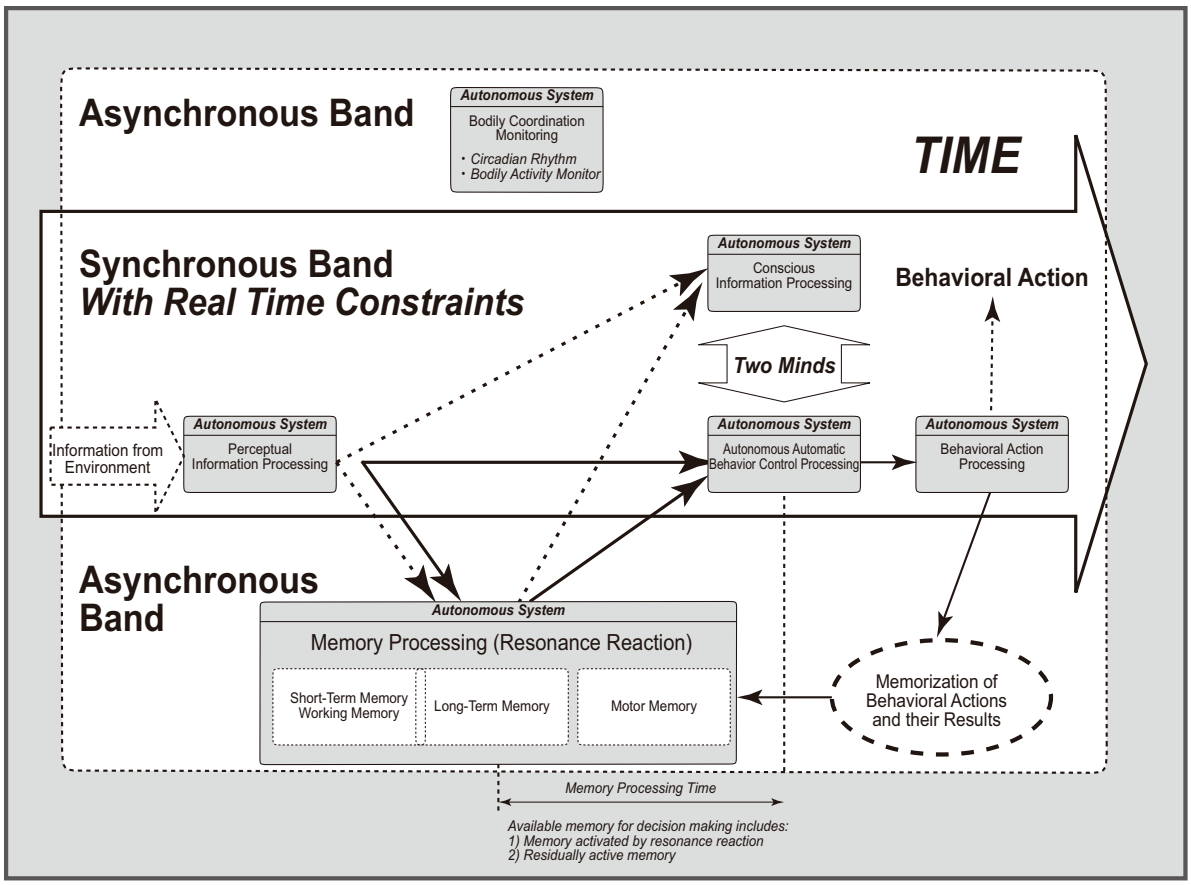


Figure 2. Outline of MHP/RT.

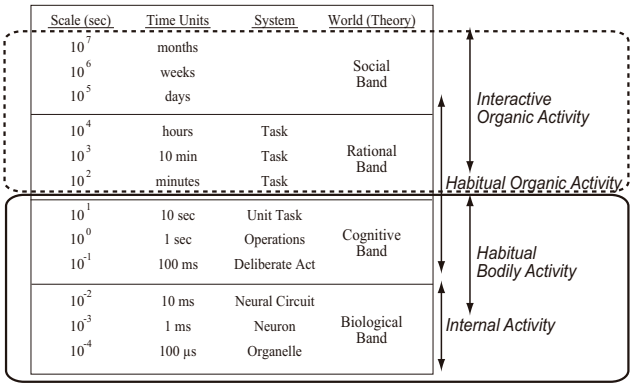


Figure 3. Newell's time scale of human action (adapted from [17]).

E. Four Processing Modes of MHP/RT

MHP/RT suggests that at a particular time *before the event*, say T_{before}, one engages in System 2's conscious processes and System 1's unconscious processes concerning the event. At a particular time *after the event*, one engages in conscious processes and unconscious processes. What one can do before and after the event is strongly constrained by the Newell's time scale of human action as shown by Figure 3. It indicates that System 2 carries out the processes surrounded by a round-cornered rectangle with dotted lines, whereas System 1 does those surrounded by a round-cornered rectangle with solid lines.

MHP/RT works under the following four processing modes, ordered from the past to the future:

- **System 2 Before Mode:** Conscious use of long-term memory before the event, i.e., System 2's operation for anticipating the future event, or decision-making.
- **System 1 Before Mode:** Unconscious use of long-term memory before the event, i.e., System 1's operation for automatic preparation for the future event, or action selection.
- **System 1 After Mode:** Unconscious use of long-term memory after the event, i.e., System 1's operation for automatic tuning of long-term memory related with the past event.
- **System 2 After Mode:** Conscious use of long-term memory after the event, i.e., System 2's operation for reflecting on the past event.

Figure 4 illustrates the four processing modes along the time dimension expanding before and after the event, which is shown as *boundary event*. At each moment, one behaves in one of the four processing modes and he/she switches among them depending on the internal and external states.

 ISBN: 978-1-61208-390-2

III. Dynamics of consciousness–emotion interaction: an explanation by NDHB-Model/RT

A. Interaction between Consciousness and Emotion

The processes in A^2BC-layer and those in C-layer are not independent. Rather, they interact with each other very intensely in some cases but very weakly in other cases. We investigate this issue in more detail below.

1) Onset of consciousness: With the onset of arousal, the sensory organs begin to collect environmental information, or paying attention to the information. This information flows into the brain, and the volume of information flow grows rapidly. As the information flow circulates in the neural networks, the center of the flow gradually emerges. It corresponds to the location where the successive firings of the neural networks concentrate. At this time, the center of information flow induces activities in C-layer via the cross-links in the neural networks.

2) Conscious activities: Figure 5 depicts the state of the brain when consciousness starts working. The location of consciousness is indicated as a dot in C-layer. In many cases, the working of consciousness includes such cognitive activities as comprehension of self-orientation and an individual's circumstances. The judgement on what decision-making is needed for the current situation is equivalent to initiating some action to move the location of consciousness to an appropriate direction. The direction of movement is determined by the information needs at that time. It could move either in the direction in which the initial information will be deepened (left in the figure, moving upstream in the information flow) or to the direction in which the initial information will be widened (right in the figure, moving downstream in the information flow). The density of information would change depending on how far the center of consciousness would have moved. However, the location of the consciousness would not move when carrying out a routine task.

Koch and Tsuchiya [18] suggested that attention and consciousness are two distinct brain processes. They showed functional roles of attention and consciousness and the four ways of processing visual events and behaviors. The upstream move of consciousness to the information flow corresponds to "attention with consciousness" in [18] and the downstream move corresponds to "consciousness in the near absence of attention."

3) Emergence of emotion: After the onset of consciousness, a new thread of information coming into the brain via the sensory organs triggers successive firing within the neural networks. This causes a new information flow in the brain that reflects the past experience that resonates with the input information. If there is a discrepancy between the new information flow (the dotted line in the figure) and the existing information flow (the solid line in the figure), emotion emerges. Emotion works to reduce the amount of discrepancy. This view is consistent with the one suggested by Tsuchiya and Adolphs [19], discussing the interaction between emotion and consciousness by reviewing experimental studies.

4) Determination of next behavior: When A^2BC-layer works continuously within its capacity, consciousness does not interfere with the working of A^2BC-layer but monitors the individual's behavior, prepares for the next behavior, and/or

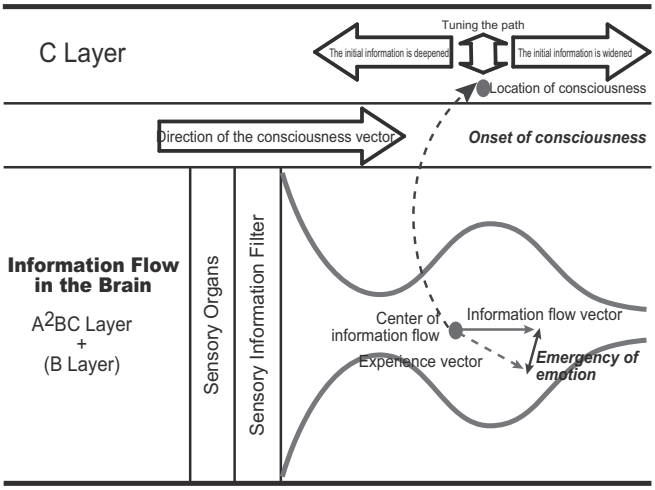


Figure 5. Onset of consciousness and emergence of emotion

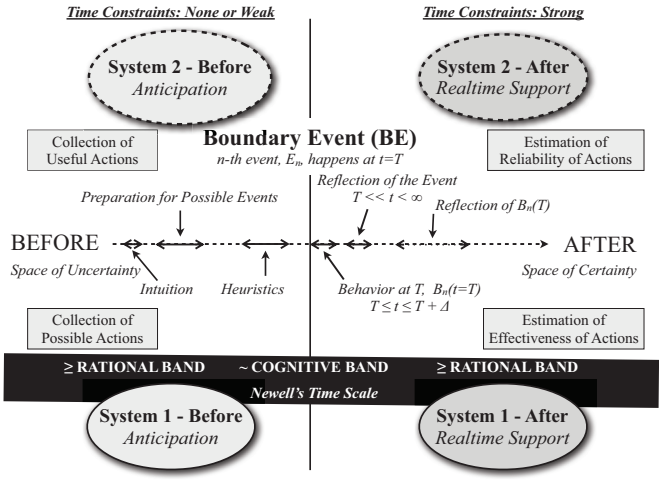


Figure 4. How the Four Processing Modes work (adapted from [9]).

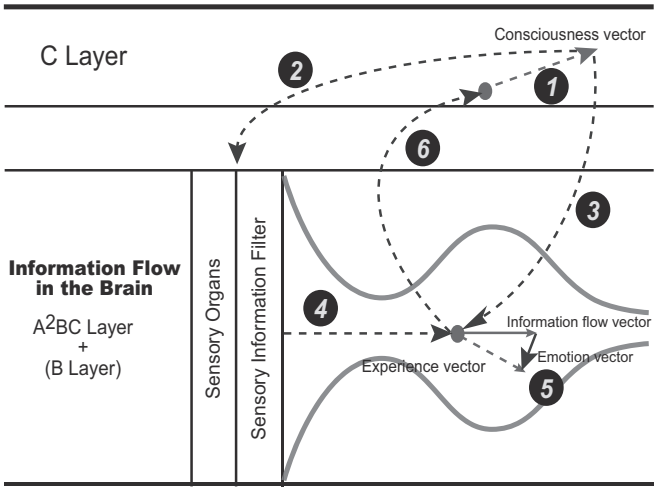


Figure 6. Determination of next behavior

ponders issues that come to mind. However, if A^2BC-layer has difficulty in determining the next behavior, C-layer takes over and determines it. However, note that decision-making deals with planning for future behavior in the "System 2 Before Mode." Actions that will be taken actually in the ever-changing real world are determined by the system flexibly in an ad-lib fashion in the "System 1 Before Mode" [9][16].

5) Summary: The following depicts the flow of the processes that would happen (see Figure 6).

1) Consciousness determines the next behavior by considering the current emotion state, which is the functional aspects of an emotion [20], and the self-recognition.
2) Consciousness tunes the orientation of the sensory organs in preparation for initiating the next behavior just determined.
3) Consciousness commands initiating the next behavior.
4) The behavior results in changes in the information flow.
5) The direction of emotion changes.
6) The new state of emotion affects the process of determining the next behavior.

B. Synchronization between the C layer and the A^2BC layer: MHP/RT's perspective

We assume that the C layer and the A^2BC layer operate together in order to determine the next behavior. However, as described above, the interaction between them could be weak or strong, depending on the situation. There thus needs to be a synchronization mechanism for them to work together appropriately. The degree of discrepancy could be measured by the amount of efforts to re-establish good synchronization between the two systems.

We suggest that the visual-frame reconstruction process in the C layer should be used for establishing synchronization between the C layer and the A^2BC layer. In Figure 2, this synchronization process is indicated schematically as an arrow with the label "Two Minds." The C layer predicts the representation of the visual frame of reference that should appear in the future and uses it for synchronization. The information flow for this process is indicated in the dotted lines in Figure 2, which occurs in the characteristic times surrounded by a round-corner rectangle with dotted lines in Figure 3. On the other hand, when the A^2BC layer mainly controls the behavior, the visual-frame updating rate would be around 10 frames per second. The information flows as indicated by the solid lines in Figure 2, with the characteristic times in the round-corner rectangle with solid lines in Figure 3.

When the C layer mainly controls the behavior as in the former case, the rate would become lower and vary depending on the interest of consciousness. In the latter case, the C layer would monitor the self-behavior by occasionally matching the expected visual frame of reference and the real visual frame of reference in the A^2BC layer. For the former situation, the visual-frame density is high but the information density is low; for the latter situation, the visual-frame density is low but the information density is high. Discrepancy would be detected easier in the former case than the latter.

IV. Relationships between Two Minds and Emotions

A. Taxonomy of emotions: behavioral perspective

Table I summarizes the relationships between Two Minds and emotions as a combination of the states of C-layer and A^2BC-layer. The top half of the table lists the kinds of decision-makings that C-layer would do before some event happens. Depending on the intensity of the signals emitted from A^2BC-layer and the self-estimate of the state of the system, C-layer decides to do something with large effort, small effort, or just do nothing, or do nothing intentionally. Note that no emotion will take place at the time when C-layer makes decisions concerning the system's future.

On the other hand, as shown in the bottom half of the table, emotions will emerge when actions are carried out by A^2BC-layer. A specific emotion type would emerge depending on the combinations of the possible states of the following four conditions:

1) the signal intensity of A^2BC-layer,
2) C-layer's estimate of the system state,
3) the nature of C-layer's decision-making, and
4) the result of A^2BC-layer's action.

For example, in Case 9, though A^2BC-layer emits good signals, C-layer estimates the situation to be fearful. It decides not to do anything. However, A^2BC-layer reacts to the situation autonomously and the situation eventually turns good. C-layer feels relieved. Feeling is the conscious experience of the emotion [21]. In sum, this table provides a taxonomy of emotions in terms of the activities of A^2BC-layer and C-layer. The architecture-based approach towards taxonomy of emotion this paper has taken is different from traditional one, which studies in detail the relationships between the observed phenomena and the triggering conditions. The strength of this approach is that it would be free from increasing complexity of society and diversity of cultures.

B. Emotion initiation via memory processes

The processes depicted in Figure 5 include activation of memory via information flow in the brain. In order to disentangle these processes, we first show Figure 7 that illustrates how each MD-frame, to be explained below, is created as the result of working of autonomous processes in MHP/RT and how MD-frames are mutually interrelated [22]. This essentially details the process "Memorization of Behavioral Actions and their Results" shown by the dotted oval in the diagram of MHP/RT, Figure 2, by considering neuronal activities that actually happen. The basic idea is that each autonomous system has its own memory.

1) MD-frame: MHP/RT assumes that memory is organized by Multi-Dimensional Frame for storing information. It is a primitive cognitive unit that conveys information that can be manipulated by brain under various constraints.

Object cognition occurs as follows [23]:

1) Collecting information from the environment via perceptual sensors;
2) Integrating and segmenting the collected information, centering on visually collected objects;
3) Continuing these processes until the necessary objects to live in the environment are obtained.

 ISBN: 978-1-61208-390-2

TABLE I. RELATIONSHIPS BETWEEN TWO MINDS AND EMOTIONS.

System 2's before-event expectation					
Case	*C-layer's before-event decision-making*	*Signasl emitted from A^2BC-layer*	*C-layer's estimate when expectation was formed*		
1	do something with small effort	stable (no signal)	relaxed		
2	do nothing intentionally	bad	prepared for bad		
3	do something with large effort	bad	positive		
4	do something with large effort	good	strongly positive		
5	do nothing	good	calm		
System 2's after-event decision-making					
Case	*C-layer after-event decision-making*	*Signals emitted from A^2BC-layer*	*C-layer's estimate when decision-making was done*	*Result of A^2BC-layer's action*	*Emotion after A^2BC-layer's action was taken*
6	do something with small effort	good	good	+	satisfaction
				−	shock, lostness
7	do nothing intentionally	bad	bad	+	amazement, pleasure
				−	regret, despair
8	do something with large effort	bad	uneasy	+	self-praise
				−	apology
9	do nothing	good	fearful	+	relief
				−	regret

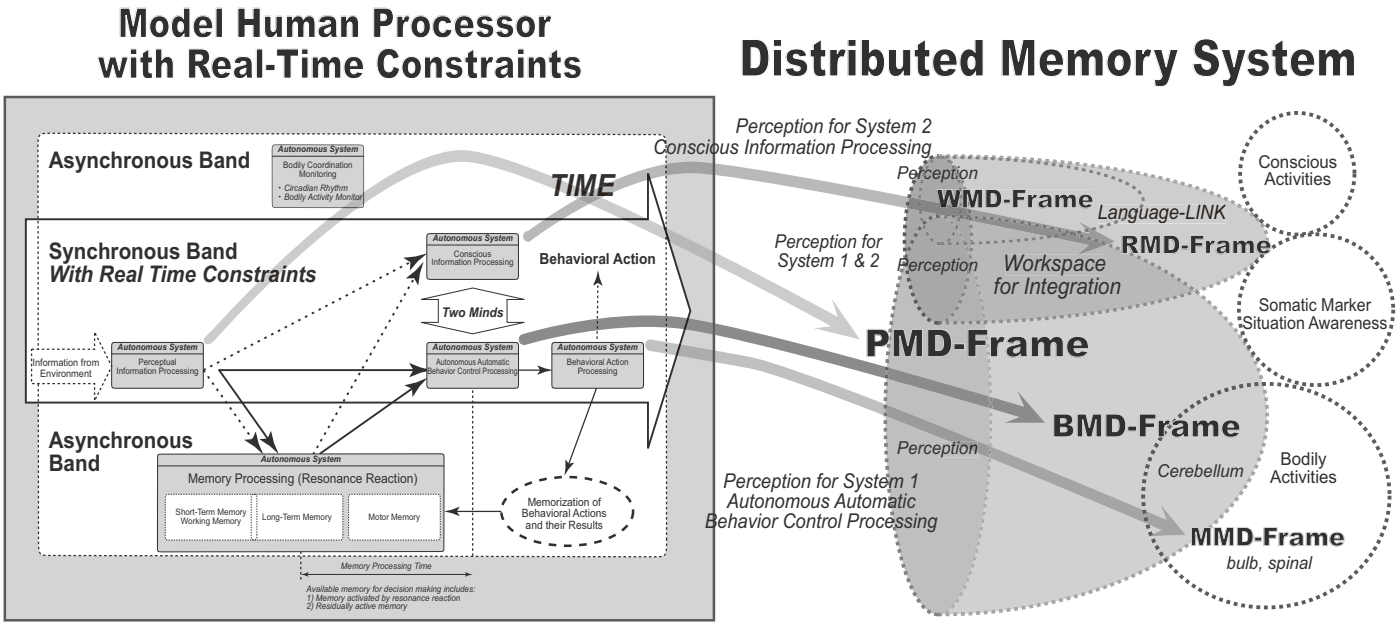


Figure 7. MHP/RT and the distributed memory system [22].

These objects are then used independently in Systems 1 and System 2, and memorized after integrating related entities associated with each system. Due to the limitation of the brain's processing capability, the range of integration is limited. Therefore, System 1 memory and System 2 memory should differ. However, they could share objects originating from perceptual sensors. Thus, when objects, that are the result of the just-finished integration and segmentation process, are processed in the next cycle, representation of the objects may serve as the common elements to combine System 1 memory and System 2 memory to form an inter-system memory. We call this memory the Multi-Dimensional (MD) -frame.

There are five kinds of MD-frame in MHP/RT.

PMD (Perceptual Multi-Dimensional)-frame constitutes perceptual memory as a relational matrix structure. It collects information from external objects followed by separating it into a variety of perceptual information, and re-collects the same information in the other situations, accumulating the information from the objects via a variety of different processes. PMD-frame incrementally grows as it creates memory from the input information and matches it against the past memory in parallel.

MMD (Motion Multi-Dimensional)-frame constitutes behavioral memory as a matrix structure. The behavioral action processing starts when unconscious autonomous behavior shows after one's birth. It gathers a variety of perceptual

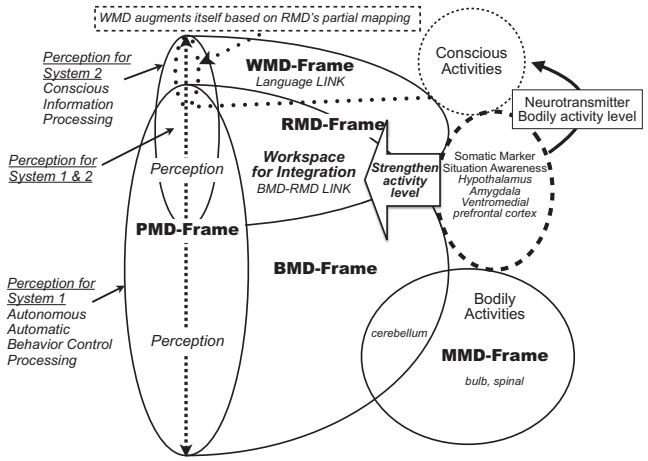


Figure 8. MD-frames and emotion.

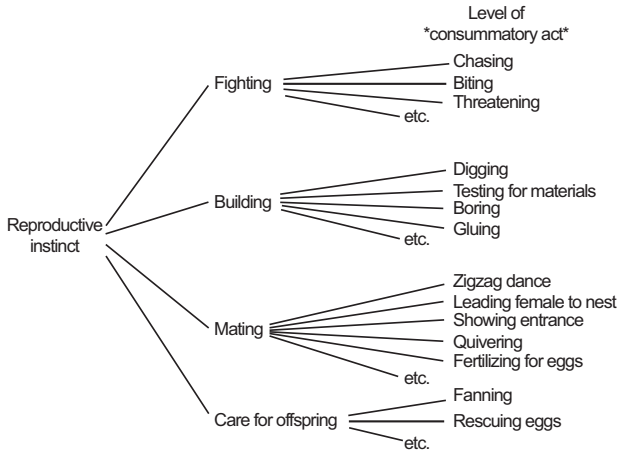


Figure 9. Fixed hierarchy in the behaviors accompanied with instinct of procreation (redrawn Figure 8.10 in [24])

information as well to connect muscles with nerves using spinals as a reflection point. In accordance with one's physical growth, it widens the range of activities the behavioral action processing can cover autonomously.

BMD (Behavior Multi-Dimensional)-frame is the memory structure associated with the autonomous automatic behavior control processing. It combines a set of MMD-frames into a manipulable unit.

RMD (Relation Multi-Dimensional)-frame is the memory structure associated with the conscious information processing. It combines a set of BMD-frames into a manipulable unit. The role BMD-frames play for RMD-frame is equivalent to the role MMD-frames play for BMD-frame.

WMD (Word Multi-Dimensional)-frame is the memory structure for language. It is constructed on a very simple one-dimensional array.

2) Consciousness/unconsciousness in MD-frames: In the MHP/RT's four processing modes, conscious processes can happen before the event and after the event. On the other hand, a series of events happen as time goes by, and each of them is processed by MHP/RT consciously or unconsciously. For the conscious processes, WMD-Frame and RMD-Frame are relevant and consciousness occurs at those levels. These frames are associated with perceptual MD-Frame. Consciousness is the phenomenon that connects WMD-Frame and RMD-Frame, and the connections are established via their relationships with BMD-Frame and MMD-Frame that are associated with the shared PMD-Frame. Considering this way, we can reach the conclusion that the phenomenon of consciousness is one aspect of the nature of memory system, or MD-frames.

Metaphorically speaking, consciousness is one of tips of icebergs that appear above the sea level, and the tips are inter-related with each other via the unseen relationships established below the sea level. A system of icebergs develops in the natural condition of seawater and atmosphere, which may or may not be trivial for any people.

3) Emotion initiation in MD-frames: Figure 8 illustrates processes in MD-frames with the focus of emotion. Memory activation originates from perception and spreads in MD-frames. In normal operation, active memory regions are used for organizing behavior. For the conscious system, most active memory regions would connect to consciousness and have effects on conscious activities. For the unconscious bodily movement system, most active memory regions corresponding to respective body parts would directly guide action selections in parallel.

Somatic markers directly guide the action selections that are carried out in a feed-forward way. On the other hand, they have indirect effects on conscious activities by providing integrated information about the current status of the body via receptors of the conscious system where neurotransmitters' local density represents the integrated response to the current status of the body. In other words, emotion corresponds to the internal activities that coordinate conscious processes and unconscious processes to work coherently in the ever-changing environment. As [21] put it, emotion emerges when consciousness is recognized for the first time. Feeling appears when the emotion is analyzed ecologically and recognized at the later time.

V. CONCLUSION

This paper suggested that emotion is a means for establishing synchronization between consciousness and unconsciousness internally via memory processes, with its taxonomy on the basis of the Four Processing Modes of MHP/RT, which is a cognitive architecture under NDHB-Model/RT with the guiding concepts of O-SCFT and O-PDP.

Figure 9 shows the fixed hierarchy in the behaviors accompanied with instinct of procreation through observation of behavior of Nemipterus virgatus [24]. Since the evolution of vertebrata is no more than history of increasing complexity, the results of any ecological analyses of emotions would tend to lead to thesaurus-like similar implicit structures, which is free from complication of society and diversity in culture, as suggested in this paper.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number 24531274.

 ISBN: 978-1-61208-390-2

REFERENCES

[1] D. Kahneman, "A perspective on judgment and choice," American Psychologist, vol. 58, no. 9, 2003, pp. 697–720.

[2] J. S. B. T. Evans and K. Frankish, Eds., In Two Minds: Dual Processes and Beyond. Oxford: Oxford University Press, 2009.

[3] J. R. Anderson and C. Lebiere, The Atomic Components of Thought. Mahwah, NJ: Lawrence Erlbaum Associates, 1998.

[4] J. R. Anderson, How can the Human Mind Occur in the Physical Universe? New York, NY: Oxford University Press, 2007.

[5] J. E. Laird, A. Newell, and P. S. Rosenbloom, "Soar: An architecture for general intelligence," Artificial Intelligence, vol. 33, 1987, pp. 1–64.

[6] W. G. Kennedy and M. Bugajuska, "Integrating Fast and Slow Cognitive Processes," in Proceedings of the International Conference on Cognitive Modeling (ICCM 2010), D. D. Sulvucci and G. Gunzelmann, Eds. Austin, TX: Cognitive Science Society, 2010, pp. 121–126.

[7] J. R. Anderson, D. Bothell, M. D. Byrne, S. Douglass, C. Lebiere, and Y. Qin, "An integrated theory of the mind," Psychological Review, vol. 111, no. 4, 2004, pp. 1036–1060.

[8] M. Kitajima and M. Toyota, "Simulating navigation behaviour based on the architecture model Model Human Processor with Real-Time Constraints (MHP/RT)," Behaviour & Information Technology, vol. 31, no. 1, 2012, pp. 41–58.

[9] ——, "Decision-making and action selection in Two Minds: An analysis based on Model Human Processor with Realtime Constraints (MHP/RT)," Biologically Inspired Cognitive Architectures, vol. 5, 2013, pp. 82–93.

[10] M. Kitajima, H. Shimada, and M. Toyota, "MSA:Maximum Satisfaction Architecture – a basis for designing intelligent autonomous agents on web 2.0," in Proceedings of the 29th Annual Conference of the Cognitive Science Society, D. S. McNamara and J. G. Trafton, Eds. Austin, TX: Cognitive Science Society, 2007, p. 1790.

[11] M. Kitajima, M. Toyota, and H. Shimada, "The Model Brain: Brain Information Hydrodynamics (BIH)," in Proceedings of the 30th Annual Conference of the Cognitive Science Society, B. C. Love, K. McRae, and V. M. Sloutsky, Eds. Austin, TX: Cognitive Science Society, 2008, p. 1453.

[12] M. Toyota, M. Kitajima, and H. Shimada, "Structured Meme Theory: How is informational inheritance maintained?" in Proceedings of the 30th Annual Conference of the Cognitive Science Society, B. C. Love, K. McRae, and V. M. Sloutsky, Eds. Austin, TX: Cognitive Science Society, 2008, p. 2288.

[13] I. Prigogine and I. Stengers, Order Out of Chaos. William Heinemann, 8 1984.

[14] J. L. McClelland, D. E. Rumelhart, and P. R. Group, Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Psychological and Biological Models (A Bradford Book). The MIT Press, 7 1987.

[15] D. Morris, The nature of happiness. London: Little Books Ltd., 2006.

[16] M. Kitajima and M. Toyota, "Four Processing Modes of *in situ* Human Behavior," in Biologically Inspired Cognitive Architectures 2011 - Proceedings of the Second Annual Meeting of the BICA Society, A. V. Samsonovich and K. R. Jóhannsdóttir, Eds. Amsterdam, The Netherlands: IOS Press, 2011, pp. 194–199.

[17] A. Newell, Unified Theories of Cognition (The William James Lectures, 1987). Cambridge, MA: Harvard University Press, 1990, page 122, Fig. 3-3.

[18] C. Koch and N. Tsuchiya, "Attention and consciousness: two distinct brain processes," Trends in Cognitive Sciences, vol. 11, no. 1, 2007, pp. 16 – 22.

[19] N. Tsuchiya and R. Adolphs, "Emotion and consciousness," Trends in Cognitive Sciences, vol. 11, no. 4, 2007, pp. 158 – 167.

[20] J. A. Lambie and A. J. Marcel, "Consciousness and the varieties of emotion experience: A theoretical framework," Psychological Review, vol. 109, 2002, pp. 219–259.

[21] A. Damasio, Feeling of What Happens: Body and Emotion in the Making of Consciousness. Orlando, FL: Houghton Mifflin Harcourt, 1999.

[22] M. Kitajima and M. Toyota, "Topological Considerations of Memory Structure," in Procedia Computer Science, BICA 2014. 5th Annual International Conference on Biologically Inspired Cognitive Architectures, vol. 99, 2014, pp. 326–331.

[23] ——, "The Role of Memory in MHP/RT: Organization, Function and Operation," in Proceedings of ICCM 2012: 11th International Conference on Cognitive Modeling, 2012, pp. 291–296.

[24] L. W. Swanson, Brain Architecture. Oxford University Press, 2011.

 ISBN: 978-1-61208-390-2

Corpus Callosum Shape Changes in Early Alzheimer's Disease: An MRI Study Using the Automatic Deformable Model

Amira Ben Rabeh, Faouzi Benzarti, Hamid Amiri
Signal, image and technology information (LR-SITI)
National Engineering School of Tunis (ENIT), Manar, Tunisia
Emails: amira.benrabeh@gmail.com benzartif@yahoo.fr hamidlamiri@yahoo.fr

Mouna Bouaziz
Radiology Service, Institut MT Kassab
Hospital Kassab, Ksar Said, Tunis,Tunisia
Email: bouaziz_mouna@yahoo.fr

***Abstract*—In this paper, we propose a solution to diagnose Alzheimer's pathology; we are interested in changing the shape of the Corpus Callosum (CC). It is the commissure of the brain and Alzheimer diseases manifests by a significant reduction of its volume. To do this, we used a classification method based on decision trees and the Active Shape Model (ASM) to extract the lesion study. For the deformable model, we added the following contribution: integration of a priori knowledge to automate the initialization of the average shape. For the pretreatment step, we used the median filter. Our method is validated by a physician to diagnose Alzheimer's disease.**

Keywords—Alzheimer; Corpus Callosum (CC); active shape model (ASM).

I. INTRODUCTION

Medical images are now ubiquitous in the clinical portion in brain imaging: anatomic imaging (Computed Tomography (CT), Magnetic Resonance Imaging (MRI)), vascular (Magnetic Resonance Angiography (MRA)) and functional (Positron Emission Tomography (PET), Single Photon Emission Computed Tomography (SPECT), Functional Magnetic Resonance Imaging (fMRI), electroencephalography (EEG), magnetoencephalography (MEG)). The amount of information increases even more when multiple images are acquired on the same patient to exploit the complementarity of different ways, or to follow a temporal evolution. Finally, these images are often accompanied by metadata on the patient's history and cerebral pathology. With all these images and complexity, the doctor can usually visually extract it as incomplete information [1]-[4].

In this work, we are particularly interested in the description of the surfaces of the Corpus Callosum and their classification to facilitate the diagnosis of Alzheimer's disease. The shape analysis and classification are part of an indivisible digital compact processing chain and automatic (or semiautomatic) Computer-Aided Diagnosis (CAD). Thus, a good evaluation of the performance of such part of description or classification requires the mastery of the entire chain of diagnosis.

Our work is organized as follows: first, we describe the previous works. In the second section, we define the Corpus Callosum. In the third section, we present our proposed method. In the fourth section, we provide results and discussion.

II. PREVIOUS WORK

There are many methods for diagnosing Alzheimer's diseases, such as:

A. VBM (Voxel Based Morphometry)

VBM is a method developed by Ashburner et al. [5] in 2000. The objective of VBM is to detect significant differences in gray matter between two groups of subjects by voxel to voxel tests. VBM comprises four steps: normalization stereotactic images on the same area, followed by segmentation of the images, and then smoothing the maps of gray matter obtained, and finally, the application of parametric statistical tests voxel to voxel.
It is a performant method but it used just the sagittal section.

B. Method Proposed by Olfa Ben Ahmed

She used the Region of Interest (ROI) to extract the hippocampus and cingulate cortex. For the classification step, it uses the Bag of Visual World (BOVW) method. This approach is applied separately on both ROI (hippocampus and cingulate cortex). The role of this model is to combine the extracted features of each ROI to build a visual vocabulary. In addition, the region differs from one projection to another [6]. Thus, we choose to perform the procedure grouping three times from different projections (sagittal, axial and coronal) and generating a visual vocabulary by projection.

C. ASM+D(Active Shape Method + Distance)

In addition to the a priori knowledge about the shape and deformation modes of the structures studied, this method consists in another acquaintance on the change in the distance between these structures. This new knowledge is estimated in a learning phase which is to be deducted from a set of sample images and based on principal component analysis, a model describing a distance variation space allowed distance between structures [7]. This controls the evolution of initial estimates during the localization phase and ensures the maintenance of the inter-pattern distance in the space allowed.

 ISBN: 978-1-61208-390-2

III. CORPUS CALLOSUM AND ALZHEIMER DISEASE

The Corpus Callosum is a commissure (through union between two parties) section of the brain. The axons beam interconnects the two cerebral hemispheres. This is the largest commissure of the brain because it connects the four lobes of the brain between them (frontal lobes, temporal, parietal and occipital left and right). The Corpus Callosum, therefore, ensures the transfer of information between the two hemispheres and coordination [8]-[13]. Other commissures are the fornix, the anterior cingulate and the commissure. The CC has been a region examined extensively for indications of Alzheimer various pathology [14]-[19]. This manifests by a significant reduction of its volume.

IV. MATERIAL

Our approach is applied to reference MRI based BRAINWEB. The images are a grayscale and size of 500 * 500 pixels. These images are made by experts.

V. PROPOSED METHOD

Our system is presented in Figure 1.

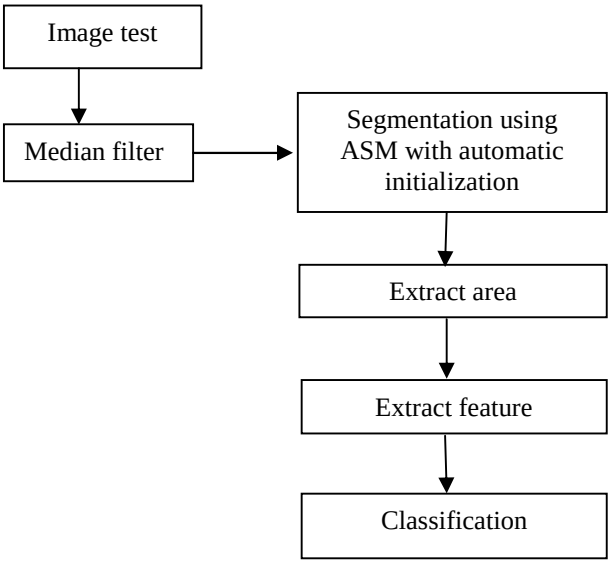


Figure.1. Diagram of the proposed method

A. Median Filter

It is specifically effective against noise pepper and salt in images to grayscale. The operation is to replace the value of a pixel with the median value of all the pixels in its neighborhood.

B. The Active Contour Model with Automatic Initialisation

Deformable models represent algorithms for segmentation images by the contour. The introduction of these models in image processing was done by Kass et al. [20] who proposed the first known as "active contours" model. Then, the idea has been widely adopted and developed to give rise to several other models depending on the particular problems and the nature of the processed images. We have grouped these models into three main classes, namely (i) parametric models whose curve is explicitly represented by a parametric function, (ii) geometric models whose curve is represented implicitly considered the border of level zero of a function, and (iii) higher order statistical models that are based on a preliminary statistical analysis of the variation structures. All these models admit an interesting property, which is the integration of a priori knowledge about the structures of interest in the process of segmentation. They have proven their performance in several application areas [20]-[23].

The ASM is a variation of deformable statistical models, which is introduced by Cootes and al. [20] in order to extract complex and non-rigid objects. The advantage of the ASM compared to other variants of deformable models is that the evolution of the curve is guided by a strong a priori knowledge about the geometry and deformation modes of the structure studied. This knowledge is represented by a statistical model which describes the authorized deformation space. The construction of such a model is done by applying a Principal Component Analysis (PCA) on a training set, which includes the various possible forms of the object [24]-[26].
The shape is defined by the following equation:

$$\upsilon = \bar{\upsilon} + P_f\, b_f \qquad (1)$$

were $\bar{\upsilon}$ the average shape, P_f matrix of the main modes of the deformation's shape and b_f a weight matrix representing the projection of the form υ in the database P_f.

The segmentation of dynamic structures in medical imaging is one of the most difficult problems that continues to preoccupy researchers. This problem arises especially in the manual initiation of the mean shape. To address this problem, we have integrated the concept of a priori knowledge to our automatic initialization to make the task of automatic segmentation and avoid manual initiation.

N feature points are positioned on the contour of the region of interest. Each form will be modeled by a vector constructed by concatenating the coordinates of points placed on the contour:

$$V_i = (X_{i1}, Y_{i1}, X_{i2}, Y_{i2}, X_{i3}, Y_{i3}, \ldots\ldots\ldots X_{in}, Y_{in}) \qquad (2)$$

where (X_{ij}, Y_{ij}) are the coordinates of point j in image i.
Our objective is to extract an average position; so, we will take a point of each vector:

$$D_i = (X_{i\,min}, Y_{i\,min}) \qquad (3)$$

where $(X_{i\,min}, Y_{i\,min})$ is a point in the vector V_i.

$$X_{i\,min} = \min(X_{ij}) \qquad (4)$$

Applying an analysis PCA, we can deduce the modes and the amplitudes of the change of position. This phase allows to

 ISBN: 978-1-61208-390-2

build the model describing the position variation range of the authorized position of the structure studied.

This model is defined by the following equation:

$$D=D_m + P_d b_d \quad (5)$$

where D_m is the average position, P_d the matrix of the variation modes of position and b_d the projection of the position D on the base P_d.

C. Extract Feature

The notion of form is very important because it allows us to identify the objects that surround us. The shape analysis is considered successful if it is used to describe objects in a manner similar to human perception of shapes. Color is used as a descriptor. We put the area of the CC segmented on a black background. After extracting the region of the CC, the surface and the standard deviation of this area will be calculated.

It is necessary to count the pixel number of a colored area.

X_{ij} is the pixel coordinates i, j.

nbl: number of rows. nbc: number of columns.

$$Surface = \sum_{i=0}^{nbl} \sum_{j=0}^{nbc} X_{ij} \quad (6)$$

The standard deviation is defined in probability and applied statistics. Statistically, it is a measure of spread data. It is defined as the square root of the variance, or equivalently as the quadratic mean of the deviations from the average. It has the same size as the random variable or the statistical variable.

If X is a random variable square-integrable, so belonging to the space $L^2(\Omega,A,P)$, standard deviation, usually denoted, is defined as the square root of the expected value of $(X-E[X])^2$:

$$\sigma_X = \sqrt{E[(X - E[X])]^2} \quad (7)$$

where E[x], the expected value of a real random variable is intuitively the value you expect to find, on average, if the same random experiment is repeated many times. It is written E[X]. If the variable X has a countable infinity of values x1, x2... with probabilities p1, p2,..., X expectancy is defined as:

$$E[X] = \sum_{i=1}^{\infty} X_i p_i \quad (8)$$

The lower the standard deviation is, the more homogeneous the study area is. Conversely, if it is more important, the area is more heterogeneous.

D. Classification

Decision trees [27] are tools for decision support. They consist of a set of rules for dividing a set of cases in homogeneous groups. Each rule involves a combination of tests on the descriptors of a case to a group. These rules are organized as a tree whose structure has the following meaning:

- Each node corresponds to the test.
- Each arc corresponds to the response test.

More generally, decision trees can handle any type of descriptor, provided a method available to group cases according to the descriptor. Since each test is applied to a single descriptor at a time, decision trees are well suited to handle heterogeneous cases.

For classification, we propose the following method: our database contains two types of images: reached and healthy. Each image is recorded by the initial vector of the contour, the surface and the standard deviation of the CC. The extraction of the region of interest is achieved by placing 30 points on the initial contour of the CC. If an image is to be tested one should determine the maximum surface for reached images and fixed as a threshold. For the standard deviation (std), we determine the least std for reaches images and fixed as a threshold deviation. For our database there were two thresholds, namely, a threshold of surface S_{surf} 3000 pixels and a threshold of standard deviation is 23 S_{dev}.

The diagram in Figure 2 shows the stages of our classification.

S_{obt}: surface obtained

Std_{obt} : standard deviation obtained

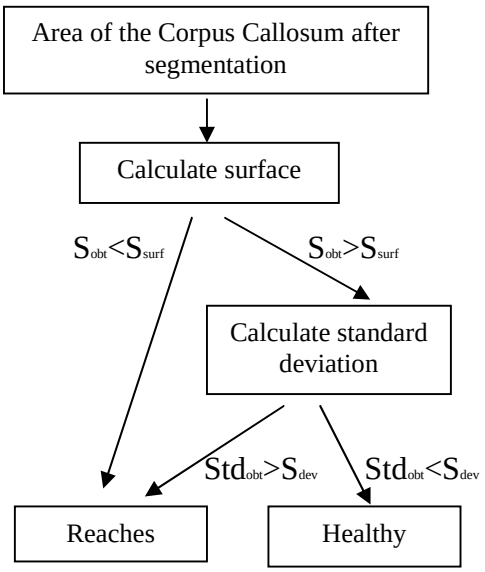


Figure.2. Diagram of the Classification algorithm

For the classification of an image after the extraction of the area of the CC, the surface of this area is determined.

- If the resulting surface is less than the threshold S_{surf} area, then the subject is directly classified as reached.
- Otherwise, the standard deviation (std) of the CC area is determined.
  - If the std exceeds the S_{dev}, then the subject is reached.
  - Otherwise, the subject is classified as normal.

 ISBN: 978-1-61208-390-2

Classification Algorithm :

```
// Step 3: Step of classification
// the surface of the Corpus Callosum Sobt was determined

        If (Sobt <Ssurf)
              Text ('Topic reached');
         else
// the standard deviation Stdobt  is determined

            If (Stdobt <Sdev)
               Text (normal Subject);
            else
               Text ('Topic reached');
```

Figure 3. Classification Algorithm

VI. RESULTS AND DISCUSSION

Segmentation is considered as the initial stage in the CAD, especially if one disregards the pretreatment stage which, according to the previous section, is not essential when processing masses. The segmentation phase is very important because the subsequent treatments (description and classification) are strongly related to the segmentation result. Indeed, a good detection of the contour of the lesion yields a true description of its characteristics. Thus, one can ensure a classification minimizing false positive rate and maximizing the rate of true negatives.

Our approach was tested on 50 images including 40 images have healthy subjects. We limit ourselves to show the results obtained for a healthy image and an image for the case reached.

Figures 4 and 5 show sample-result MRI localization obtained from a healthy and reached subject.

The first row shows the image and the original contour and the second line shows the results obtained with the application of the ASM model.

After the execution of the classification algorithm, we find the value 3000 pixels as the threshold surface and 23 as the threshold deviation.

Table I gives the values for the two studied subjects as well as the classification comparing the surface and standard deviation by the thresholds.

Subject 1:

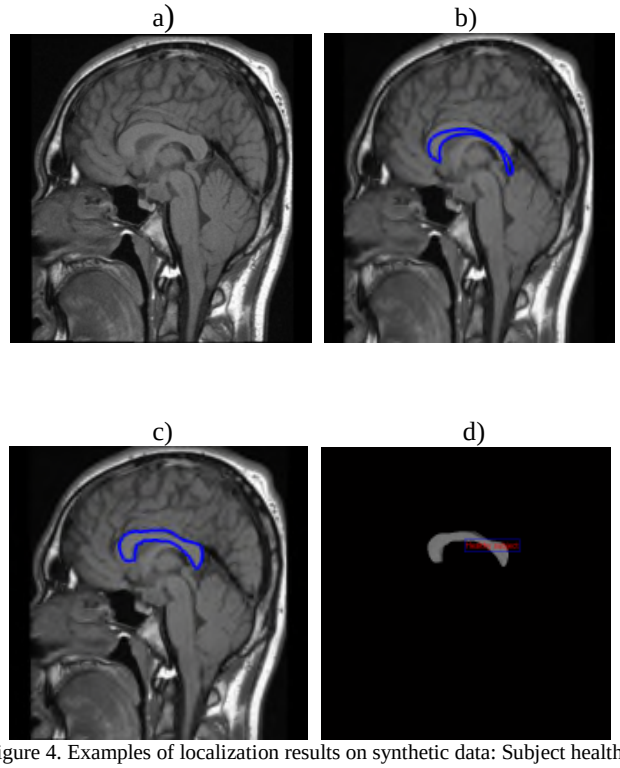


Figure 4. Examples of localization results on synthetic data: Subject healthy. (a) Original image (b) Original contour (c) Final contour (d) region detected

Subject 2:

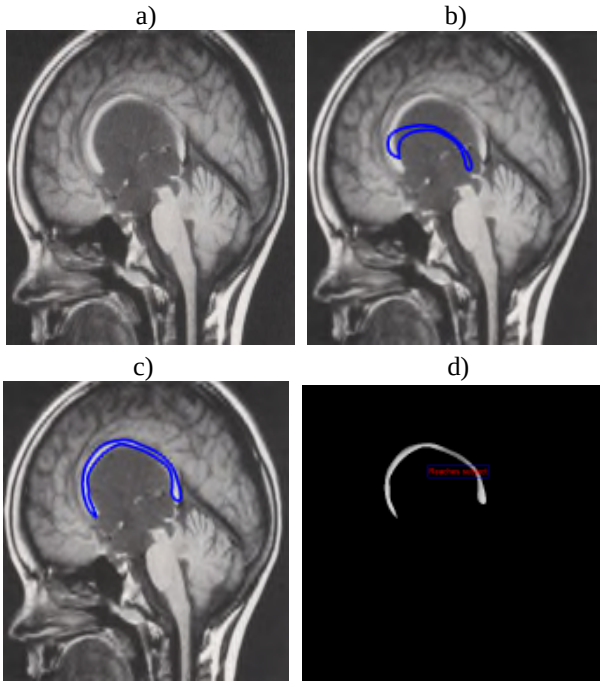


Figure 5. Examples of localization results on synthetic data: Subject reached. (a) Original image (b) Original contour (c) Final contour (d) region detected

 ISBN: 978-1-61208-390-2

TABLE I. VALUES OF THE SURFACE AND THE STANDARD DEVIATION

	Surface CC (pixel)	Standard Deviation
Subject 1 Healthy Subject	3120	13.1
Subject 2 Reaches Subject	2450	-

The surface of the CC for subject 1 is equal to 3120 pixels; this number is above the threshold surface so we move to calculate the standard deviation, which is equal to 13.1, and this is less than the standard deviation threshold therefore this subject is classified as healthy. The CC surface for subject 2 is equal to 2450 pixels is less than the threshold surface so this is classified directly as reached.

Using the calculation of standard deviation, we can analyze the region of the CC in a more robust manner. Our method can be applied on both sides of the brain. If one of the two parties does not satisfy the conditions, then we are in the case of the reached subject. The advantage of ASM is that the evolution of the curve is guided by a strong a priori knowledge about the geometry and deformation modes of the studied structure. This knowledge is represented by a statistical model which describes the authorized deformation.The construction of such model is done by applying a principal component analysis on a training set, which includes the various possible forms of the object. This model converges with the desired contours and the classification method has given us effective results. The results suggest that coprus callosum shape may be a more sensitive marker than its size for monitoring the evolution of Alzheimer disease.

We have appied the two supervised classification (k-nearest neighbors KNN [28] and decision tree [29]) methods with 50 images (40 normal and 10 AD). The KNN algorithm is among the simplest machine learning algorithms.

TABLE 2. RESULT CLASSIFICATION USING THE 2 METHODS

	TR	TFP	TFN
Decision Tree	39	5	6
KNN	35	6	9

We see that the decision tree is better than the KNN.

IV. CONCULSION AND FUTURE WORK

We proposed an approach to detect Alzheimer's diseases, based on deformable model for segmentation of the Corpus Callosum. In a first step, we used the median filter to avoid noise in the image. Then, we added a priori knowledge to the manual initialization of the mean shape to make this step automatic. After extraction of the Corpus Callosum, we used a classification method that brings in a decision tree. As perspective for this work, we propose to use a method of supervised classification as SVM (Support Vector Method).

REFERENCES

[1] Bosc, M., Heitz, F., Armspach, J.-P., Namer, I., Gounot, D. and Rumbach, L.: Automatic change detection in multimodal serial MRI: application to multiple sclerosis lesion evolution, 2003, NeuroImage 20, 643–656

[2] Radke, R.J., Andra, S., Al-Kofahi, O. and Roysam, B.: Image change detection algorithms: a systematic survey. IEEE Trans, 2005, Image Process 14(3), 294–307

[3] Bosc, M., Heitz, F., Armspach, J.-P., Namer, I., Gounot, D. and Rumbach, L.: Automatic change detection in multimodal serial MRI: application to multiple sclerosis lesion evolution. 2003, NeuroImage 20, 643–656

[4] Rey, D., Subsol, G., Delingette, H. and Ayache, N.: Automatic detection and segmentation of evolving processes in 3D medical images: Application to multiple sclerosis. In: Kuba, A., S´amal, M., Todd-Pokropek, A. (eds.) IPMI 1999. LNCS, Springer, Heidelberg 1999, vol. 1613, pp. 154–157.

[5] Ashburner J and Friston KJ. Voxel-based morphometry the methods. Neuroimage. 2000 Jun; 11(6 Pt 1):805-21.

[6] O. Ben Ahmed, J. Benois-Pineau, C Ben Amar, M. Allard and G. Catheline , Early Alzheimer disease detection with bag-of-visual-words and hybrid fusion on structural brain mri, Content-Based Multimedia Indexing'2013, Veszprêm: Hungary 2013.

[7] S. Ettaieb, K. Hamrouni and S. Ruan. Statistical modelsof shape and spatial relation-application to hippocampus segmentation. 9th International Conference on Computer Vision Theory and Applications (Visapp), Lisbon- Portugal, Janvier 2014.

[8] Cabezas M, Oliver A, Llado´ X, Freixenet J and Cuadra MB (2011) A review of atlas-based segmentation for magnetic resonance brain images. Comput Methods Programs Biomed 104:e158–e177

[9] Di Paola M, Di Iulio F, Cherubini A, Blundo C, Casini AR,Sancesario G, Passafiume D, Caltagirone C and Spalletta G (2010) When, where, and how the corpus callosum changes in MCI and AD: a multimodal MRI study. Neurology 74:1136–1142

[10] Di Paola M, Luders E, Di Iulio F, Cherubini A and Passafiume D,Thompson PM, Caltagirone C, Toga AW and Spalletta G (2010) Callosal atrophy in mild cognitive impairment and Alzheimer's disease: different effects in different stages. Neuroimage 49:141–149

[11] Di Paola M, Spalletta G and Caltagirone C (2010) In vivo structural neuroanatomy of corpus callosum in Alzheimer's disease and mild cognitive impairment using different MRI techniques: a review. J Alzheimers Dis 20:67–95

[12] Frederiksen KS, Garde E, Skimminge A, Ryberg C, Rostrup E, Baare´WF, Siebner HR, Hejl AM, Leffers AM and Waldemar G (2011) Corpus callosum atrophy in patients with mild Alzheimer's disease. Neurodegener Dis 8:476–482

[13] Hampel H, Teipel SJ, Alexander GE, Horwitz B, Teichberg D and Schapiro MB, Rapoport SI (1998) Corpus callosum atrophy is a possible indicator of region- and cell type-specific neuronal degeneration in Alzheimer disease: a magnetic resonance imaging analysis. Arch Neurol 55:193–198

[14] Hampel H, Teipel SJ, Alexander GE, Horwitz B, Teichberg D and Schapiro MB, Rapoport SI (1998) Corpus callosum atrophy is a possible indicator of region- and cell type-specific neuronal degeneration in Alzheimer disease: a magnetic resonance imaging analysis. Arch Neurol 55:193–198

[15] Teipel SJ, Bayer W, Alexander GE, Bokde AL, Zebuhr Y, Teichberg D, Mu¨ller-Spahn F, Schapiro MB, Mo¨ller HJ, Rapoport SI and Hampel H (2003) Regional pattern of hippocampus and corpus callosum atrophy in Alzheimer's disease in relation to dementia severity: evidence for early neocortical degeneration. Neurobiol Aging 24:85–94

[16] Thomann PA, Wustenberg T, Pantel J, Essig M and Schroder J (2006) Structural changes of the corpus callosum in mild cognitive impairment and Alzheimer'sdisease. Dement Geriatr Cogn Disord 21:215–220

[17] Wang PJ, Saykin AJ, Flashman LA, Wishart HA, Rabin LA, Santulli RB, McHugh TL and MacDonald JW, Mamourian AC (2006) Regionally specific atrophy of the corpus callosum in AD, MCI and cognitive complaints. Neurobiol Aging 27:1613–1617

[18] Weis S, Jellinger K and Wenger E (1991) Morphology of the corpus callosum in normal aging and Alzheimer's disease. J Neural Transm Suppl 33:35–38

 ISBN: 978-1-61208-390-2

[19] Zhu M, Gao W, Wang X, Shi C and Lin Z (2012) Progression of corpus callosum atrophy in early stage of Alzheimer's disease: MRI based study. Acad Radiol 19:512–517.

[20] T. Cootes, C. Taylor, D. Cooper and J. Graham. Active Shape Models – Their Training and Application. Computer Vision and Image Understanding, January 1995,vol. 61, No. 1, pp 38-59.

[21] S. Ettaieb, N. Khlifa and K. Hamrouni. Follow up of the left ventricle movement in dynamic scintigraphic images based on a spatio-temporal priori knowledge.4th International Symposium on Image/Video Communications over fixed and mobile networks. ISIVC 2008. Bilbao Spain. July 9-11, 2008

[22] H. Ghassan and T Gustavsson. Combining Snakes and Active Shape Models for Segmenting the Human Left Ventricle in Echocardiography Images. IEEE Computers in Cardiology, 2000, Vol 27.

[23] P. He and J. Zheng. 'Segmentation of tibia bone in Ultrasound images using Active Shape Models'.23rd Annual Conference– IEEE/EMBS, Istanbul,TURKEY. Oct 2001

[24] T. Cootes, A. Hill, C. Taylor, and J. Haslam. The use of active shape models for locating structures in medical images. Image and Vision Computing 12(6):355- 366. 1994

[25] Z. zheng and F. yang . Enhanced active shape model for facial feature localization Proceedings of the Seventh International Conference on Machine Learning and Kunming, 12-15 July 2008.

[26] Matthias Seise, Stephen J. McKenna and Carlos A. Wigderowitz. Learning Active Shape Models for Bifurcating Contours. IEEE transactions on medical imaging, may 2007, vol. 26, no. 5.

[27] L. Breiman, J. Friedman, R. Olshen and C. Stone, Classification and regression trees, Wadsworth & Brooks, 1984.

[28] Hechenbichler, K. and Schliep K. (2004) Weighted k-nearest-neighbor techniques and ordinal classification. Sonderforschungsbereich 386, paper 399.

[29] L. Breiman, J. Friedman, R. Olshen and C. Stone, Classification and regression trees, Wadsworth & Brooks, 1984.

 ISBN: 978-1-61208-390-2

Connectome Pathways in Parkinson's Disease Patients with Deep Brain Stimulators

Connectome Pathways in Parkinson's Disease Patients with DBS

Giorgio Bonmassar and Nikos Makris
AA. Martinos Center for Biomedical Imaging, MGH, Harvard Medical School,
Boston, USA.
e-mails: gbonmassar@partners.org and nmakris@partners.org

Abstract— The overarching goal of this paper is to optimally control the implanted programmable generator (IPG), a critical device that delivers electrical currents or potentials to treat neurological symptoms in patients with Parkinson' Disease (PD). Current IPG programming is based on trial and error empirical assessment, which makes the treatment implementation cumbersome, long, frustrating and expensive for the patient. Furthermore, the manifestation of the effects of IPG programming in some patient populations (e.g., dystonia) can be apparent after days, weeks or even months, which makes the trial and error approach unmanageable. Thus, the optimal IPG programming is critical to alleviate the patient's neurological symptoms. Their programming relies on parameter definition, such as electrode pair, amplitude and frequency. The positioning of the electrodes in a specific anatomical target region of interest (ROI), such as STN is limited by the surgical procedure and presurgical planning. Knowledge of the morphometry of the target ROI and its topographic relationships with surrounding anatomical structures such as whiter matter fibers (such as the internal capsule and the H1 and H2 fields of Forel) and other gray matter structures (such as the zona incerta, the substantia nigra and the red nucleus) allows the precise positioning of the stimulating electrode pair. We show that connectome imaging technology provides the necessary detailed and comprehensive in-vivo imaging of white matter fiber architecture. Based on our previous DBS studies, we present a novel numerical head model to develop a novel IPG programmer to assist neurologist in the patient management.

Keywords-deep brain brain stimulation; DBS; Parkinson's Disease; PD; MRI; DTI; implanted programmable generator; IPG; programming; connectome; STN

I. INTRODUCTION

More than 100,000 Parkinson' Disease (PD) patients worldwide have been treated with Deep Brain Stimulation (DBS) during the last twenty years resulting in 25–75% improvement of movement disorder symptoms of PD. The outcome of DBS neurosurgery depends principally on the precision of the implanted electrode placement and the ability to find the optimum settings for the Implantable Pulse Generator (IPG), a critical device for post-operative clinical management [1]-[3]. Although the theoretical basis of DBS for PD targets, such as the subthalamic nucleus (STN) and the internal part of globus pallidus (GPi) has been studied extensively in the 1980s and early 1990s, the exact mechanism of how electrical stimulation affects brain cells is not known with certainty. Given that in PD loss of dopaminergic cells leads to excessive activity in the STN and GPi, it is thought that IPGs correct this abnormal activity by injecting high-frequency electrical pulses. With respect to neurosurgery per se, DBS involves minimal permanent brain changes, however, there can be side effects, which are variable. *Most common side effects with STN implants are ataxic gait and tonic muscle contractions, paresthesias and diplopia, as well as behavioral manifestations such as depression, mania and impulse dysregulation. These are thought to be related mainly to (a) misplacement of the implanted electrodes, (b) local deformation of tissue and tissue scaring due to surgery, and (c) suboptimal programming of the IPG.* Currently, there is no realistic DBS model for IPG programming taken from actual patients with DBS implant as proposed herein. The state-of-the-art numerical DBS modeling is based on a wire or set of wires, which represents the virtual DBS implant, superimposed to healthy human brains [1, 2] (IARIA). Such models are incomplete since they do not take in consideration an accurate anatomical modeling of the fine-grained composition of the anatomic structures and their surrounding architecture involved in the stimulation, anisotropic dielectric constants and the tissue scar from the surgery. Therefore, for an enhanced VPS model and IPG programming we need detailed knowledge about the i) structural and functional anatomy of the targets and surrounding tissue, ii) encapsulating tissue around the electrode and iii) conductivity and permittivity along x, y, z, s. Post-surgical management may last several years after surgery and can be difficult. Easily adjustable or programmable IPGs (with no need of further surgery) are extremely helpful and, probably the most important tools for the neurologist to manage the PD patient long term. Therefore, safe and successful use of DBS relies heavily upon our capability to program (or adjust) the IPG. Usually, in most patients there is a reduction in levodopa medication after DBS surgery of the STN. If IPG programing is optimal, besides minimizing side effects and a safer use of DBS in PD, it can reduce or discontinue pharmacological treatment in post-surgical patients with PD. However, optimal IPG programming needs detailed knowledge of anatomical structure and function of the brain circuitry involved. Thus it seems plausible that developing and testing new IPGs optimized by Virtual Patient Stimulator (VPS) models, which are informed by precise anatomical and physiological data in

 ISBN: 978-1-61208-390-2

the individual patient will improve the therapeutic effects of neurostimulation on brain circuitry and the brain structures affected in PD. In section II we introduce: (A) the MRI acquisition methods used, (B) the numerical electromagnetic simulations performed. In section III we outlined the tractography and electromagnetic fields estimated by our simulations as well as the predicted neuronal firing frequency. In section IV we discussed our results and presented the conclusion of our study.

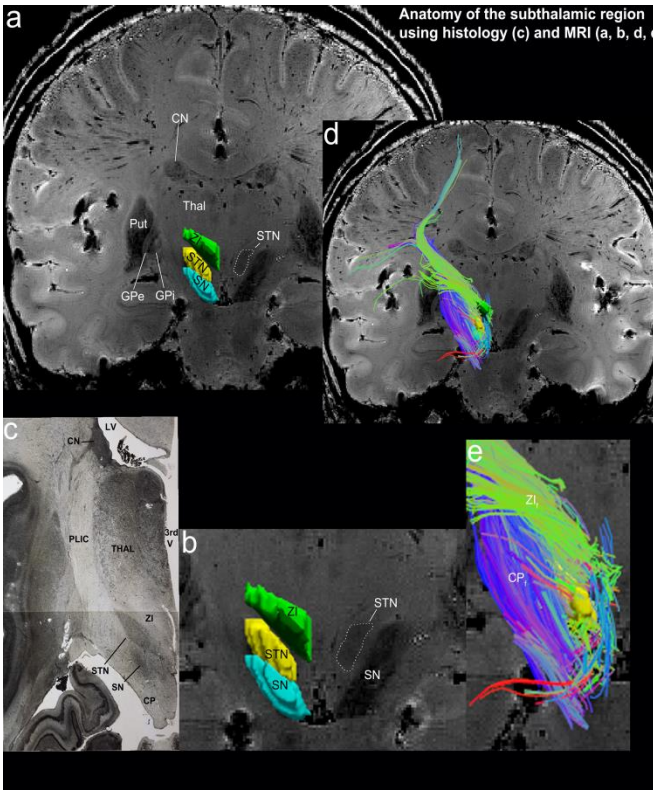


Figure 1. Anatomy of the subthalamic region using histology (c) and MRI (a, b, d, e).

II. METHODS

A. Overview of study design

The study is based on *ex-vivo* analysis of collected 7 Tesla (7T) T2* and Connectome diffusion MRI data.

(a) *Structural T2* 7 Tesla MRI* This *ex-vivo* brain consisted of MRI data of a hemisphere fixed in Periodate-Lysine-Paraformaldehyde (PLP) using the following parameters: T2*-W, 100 μm^3 isotropic resolution, TR/TE/flip=40ms/20ms/20°, 1600×1100×896 matrix. *(b)* *Connectome DSI data* are transformed into DTI data to estimate complex relative permittivity tensor $\hat{\varepsilon}^* = \frac{\varepsilon^*}{d} \cdot \mathbf{D}$, where ε* is the tissue complex relative permittivity **D** is the DTI tensor and *d* is the diffusivity. The vmPFC-BG tract is delineated using diffusion Connectome data as shown in Figure 1(b). DTI/DSI data are visually validated by comparing the computed fiber tracks with anatomical atlases with particular emphasis to the basal ganglia region. Finally, DTI/DSI data provide detailed information on the fiber tract connectivity between the STN and other parenchymal areas that is useful for DBS programming [3] and basic neuroscience research. The diffusion data set was collected on the MGH Connectome scanner with a diffusion weighted spin echo EPI sequence (1.5^3 mm^3 resolution, 140^2 matrix, FoV 210 mm, 95 slices, 128 diffusion directions at 5000 s/mm^2 and 10 b=0 acquisitions, TR 8.8 s, TE 57 ms, 3x GRAPPA acceleration, 64 channel head array) and the T1-weighted anatomical was collected on the MGH 7 T scanner with a 3D FLASH (0.4^3 mm^3 resolution, 512^2 matrix, FoV 205 mm, 352 slices, FA 30°, TR 35 ms, TE 10.2 ms, 32 channel head array).

One dataset was obtained from the Human Connectome Project (Washington Univ-Minnesota), with high spatial resolution of 1.25 mm (isotropic) (highest b-value of 3000 s/mm^2), on T1-weighted anatomical images. The *Washington U-Minnesota datasets* diffusion dataset was collected with a diffusion weighted spin echo EPI sequence (1.25^3 mm^3 resolution, 168^2 matrix, FoV 210 mm, 111 slices, 89/90/91 diffusion directions at each of 1000 s/mm^2, 2000 s/mm^2 and 3000 s/mm^2 collected with LR and RL phase encoding and 6 b=0 acquisitions, FA 78°/160°, TR 5.52 s, TE 89.5 ms, 6/8 partial Fourier, 3x multiband acceleration, 1488 Hz/px, 32 channel head array) and the T1-weighted anatomical was collected with a 3D MPRAGE (0.7^3 mm^3 resolution, 320^2 matrix, FoV 224 mm, 256 slices, FA 8° non-selective water excitation, TR 2.4 s, TE 2.14 ms, TI 1 s, asymmetric echo, 2x GRAPPA acceleration, 210 Hz/px, 32 channel head array).

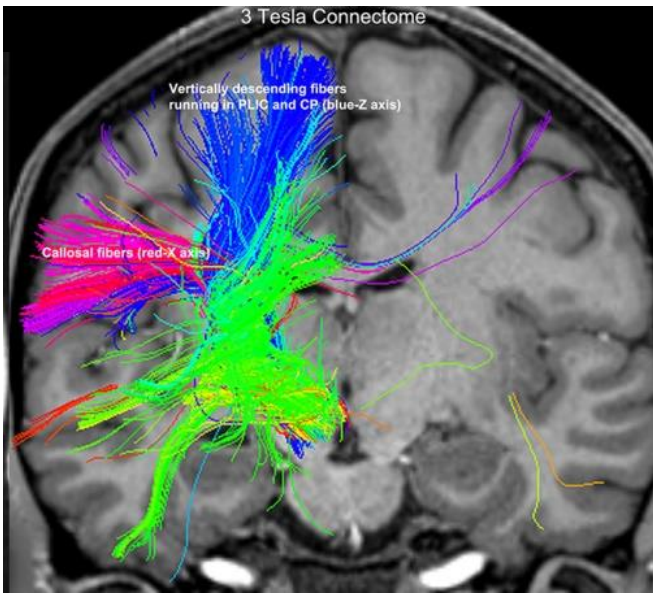


Figure 2. Diffusion imaging tractographic results in the 3 Tesla Washington U-Minnesota datasets.

B. Numerical Model of Deep Brain Stimulation implant

One or two bilateral implants, as shown by the post-operative MRI, are modeled as insulated wire(s) connected to the left and/or right targets in the head [4]. The wires are modeled as a Perfect Electrical Conductor and the dielectric is modeled as Teflon. A four-electrode connection [3] and the scar tissue are modeled in full detail reaping the benefits of the proposed 100 μm^3 isotropic resolution based on the actual Medtronic electrode set that will be used. The four electrodes are modeled as PEC and the scar tissue is modeled with the known dielectric properties.

III. RESULTS

In the following, we present preliminary results for structural imaging.

Figure 1 shows diffusion tractography results on a T1 anatomical image. We performed whole brain tractography using our state-of-the art tractography algorithm [5] capable of handling multiple fiber crossings. Subsequently, tracts connecting the STN and other brain regions were extracted. As seen in Figure 2, shows a zoomed-in portion of the tracts connecting the STN, including the ansa lentricularis fibers (green) and the other fiber bundles (blue, red). Thus, high spatial resolution, as in the connectome data, is critical for accurate tracing of these tracts, which can lead to a better understanding of the neural fiber bundles connecting STN and other sub-cortical regions around the STN. Note that, a slight misplacement of the electrode could result in the excitation of a completely different neural network, resulting in unwanted side effects. Figure 3.A shows the spatial distribution of electric field amplitude overlaid with the precise anatomy of the area surrounding the DBS implant. The field produced for the narrow bipolar stimulation configuration 1-2 is shown for three different encapsulation layer conditions. There were no differences in electric field calculated at the baseline (no encapsulation) or at the chronic stage, with a peak intensity of the electric field in both cases equal to 7.54 V/mm. Conversely, the electric field changed dramatically for the acute stage, where the higher conductivity of CSF generated a peak electric field equal to 3.59 V/mm, i.e., less than half compared to baseline or chronic stage. Furthermore, the model at the acute stage resulted in an electric field that was more spread along the electrode and asymmetric compared to the chronic stage. The electric field for the acute-stage model was characterized by higher intensity at 10mm distance from the electrode (20.33 mV/mm vs. 3.97 mV/mm). Differences in electric field between acute vs. chronic model (Figure 3.B) were also visible in the vicinity of the electrode (left), on the cortex (middle) and on the scalp (right). The stimulation was attenuated by the encapsulation, with 18% reduction in electric field amplitude delivered in the acute case and only 14% in the chronic case. The electromagnetic solution analysis was performed both in the area that surrounds the electrode and far from the electrode, i.e., on the scalp. We estimated the potential distribution for the electrode configuration 1-2 in the baseline case and compared our results with those reported in the literature. We report for the potential a drop of 84% within 4 mm of the electrode (Figure 3.C), which is in agreement with the results provided in [6]-[9] for bipolar DBS with ± 1 V voltage.

In Figure 3.D is shown the typical output of the neuron model with 117 neurons/axons, which is the status either one or zero indicates the absence or presence of an action potential. The neuron modelled was the same done by McIntyre [6], in the modeling of DBS. The electrical parameters were: conductivity=0.7/Ohm-m, membrane capacitance=0.1 uF/cm^2/lamella membrane, and membrane conductivity=0.001 S/cm^2/lamella membrane. The pulse parameters were: width = 0.1 //ms, amplitude= 3V.

IV. DISCUSSION

In order to avoid misplacements of the implanted electrodes, as well as to accurately target the sensorimotor parts of the nucleus at its dorsolateral zone [7], we need to have an accurate mapping and clear anatomical understanding of the subthalamic region. Although the allowed margin of error is 5 mm, in excellent neurosurgical procedures the error does not exceed 1 mm. However, in routine DBS practice, the margin of error is variable and several times misplacement of electrodes is such that the STN can be entirely missed. Thus the use of atlases with an estimate of the intersubject variability like the ones proposed in this study would be very useful in routine DBS neurosurgery.

Accurate tracing of the fiber connections from the STN and surrounding sub-cortical regions (e.g., substantia nigra) is critical for understanding the effect of stimulation on the neural fiber bundles connected to the STN. A small misplacement of the electrode (by a few millimeters) can result in excitation of a completely different neural circuit in the brain. Thus, accurate localization of the STN and the surrounding subcortical structures in the diffusion MRI (dMRI) images along with tractography of the associated fiber network is an essential component of our proposed work. Since the subcortical structures of interest are very small (only a few millimeters), high spatial resolution of the dMRI images is extremely important to accurately delineate these structures.

The next step will include to validate the model at a patient level. The model outlined in this abstract will be tested to check if it provides valid prediction of some or all side effects recorded in the patient and surfaced during IPG programming at different parameter settings.

V. ACKNOWLEDGMENT

National Institute of Health, National Institute of Biomedical Imaging and Bioengineering 1R21EB016449-01A1.

 ISBN: 978-1-61208-390-2

REFERENCES

[1] E. B. Montgomery, Deep brain stimulation programming : principles and practice. Oxford UK ; New York: Oxford University Press, 2010.

[2] C. C. McIntyre, S. Miocinovic, and C. R. Butson, "Computational analysis of deep brain stimulation," Expert Rev Med Devices, vol. 4, Sep 2007. pp. 615-22.

[3] C. C. McIntyre, S. Mori, D. L. Sherman, N. V. Thakor, and J. L. Vitek, "Electric field and stimulating influence generated by deep brain stimulation of the subthalamic nucleus," Clin Neurophysiol, vol. 115, Mar 2004. pp. 589-95.

[4] L. Angelone, J. Ahveninen, J. Belliveau, and G. Bonmassar, "Analysis of the Role of Lead Resistivity in Specific Absorption Rate for Deep Brain Stimulator Leads at 3 T MRI," IEEE Trans Med Imaging, Mar 22 2010. pp. 1029-38.

[5] S. A. Mohsin, N. M. Sheikh, and U. Saeed, "MRI-induced heating of deep brain stimulation leads," Phys Med Biol, vol. 53, Oct 21 2008. pp. 5745-56

[6] T. Eichele, et al. "Prediction of human errors by maladaptive changes in event-related brain networks," Proc Natl Acad Sci U S A, vol. 105, , Apr 22 2008, pp. 6173-8.

[7] J. M. Henderson, J. Tkach, M. Phillips, K. Baker, F. G. Shellock, and A. R. Rezai, "Permanent neurological deficit related to magnetic resonance imaging in a patient with implanted deep brain stimulation electrodes for Parkinson's disease: case report," Neurosurgery, vol. 57,discussion E1063, Nov 2005. p. E1063

[8] Y. Rathi, B. Gagoski, K. Setsompop, O. Michailovich, P. E. Grant, and C. F. Westin, "Diffusion propagator estimation from sparse measurements in a tractography framework," Med Image Comput Comput Assist Interv, vol. 16, 2013. pp. 510-7.

[9] N. Yousif, R. Bayford, P. G. Bain, and X. Liu, "The peri-electrode space is a significant element of the electrode-brain interface in deep brain stimulation: a computational study," Brain Res Bull, vol. 74, Oct 19 2007. pp. 361-8,

[10] S. Miocinovic, et al. , "Computational analysis of subthalamic nucleus and lenticular fasciculus activation during therapeutic deep brain stimulation," J Neurophysiol, vol. 96, Sep 2006. pp. 1569-80.

[11] A. M. Kuncel and W. M. Grill, "Selection of stimulus parameters for deep brain stimulation," Clin Neurophysiol, vol. 115, Nov 2004. pp. 2431-41.

[12] C. C. McIntyre, A. G. Richardson, and W. M. Grill, "Modeling the excitability of mammalian nerve fibers: influence of afterpotentials on the recovery cycle," J Neurophysiol, vol. 87, Feb 2002. pp. 995-1006.

[13] B. A. Strickland, J. Jimenez-Shahed, J. Jankovic, and A. Viswanathan, "Radiofrequency lesioning through deep brain stimulation electrodes: a pilot study of lesion geometry and temperature characteristics," J Clin Neurosci, vol. 20, Dec 2013. pp. 1709-12.

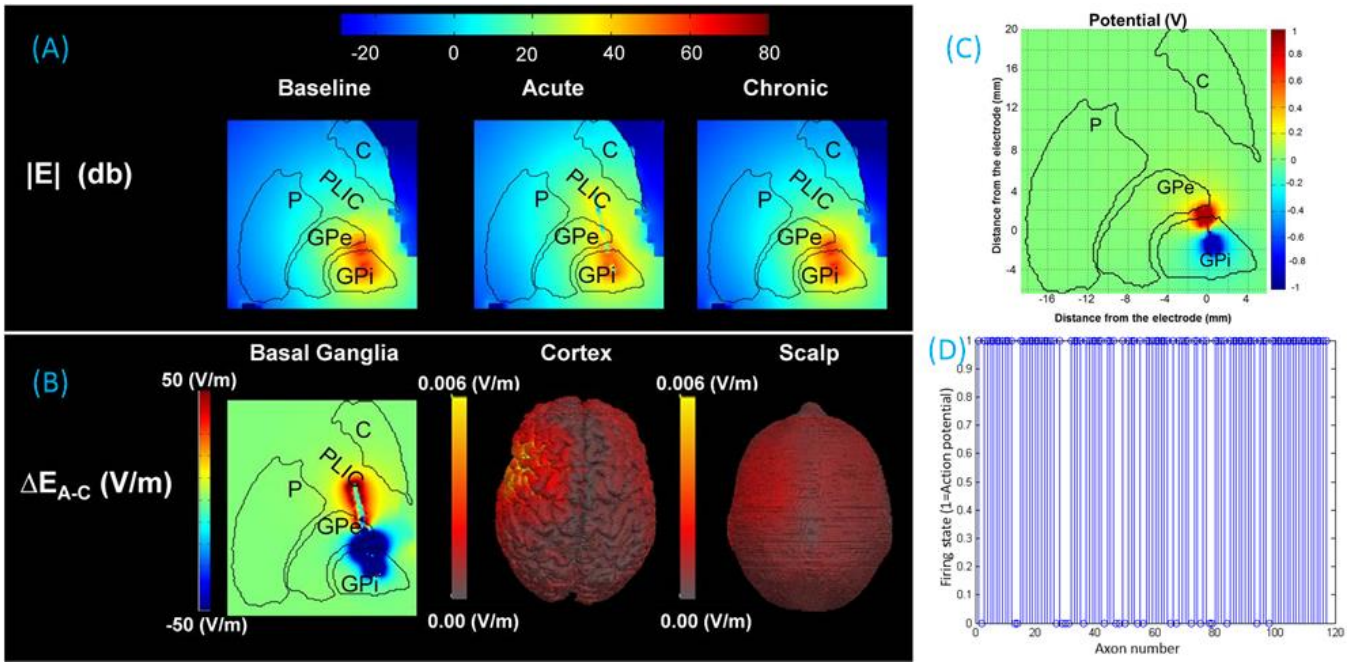


Figure 3. Electromagnetic and neural simulations. (A) Electric field |E| for: baseline (left), acute (middle), and chronic (right). (B) ΔEA-C between the acute and the chronic stage (left), on the cortex (middle row), and on the scalp (right) (C) Distribution of the potential in the vicinity of the electrode. (D) Example of firing state.

 ISBN: 978-1-61208-390-2

Proposal of an Educational Design to Improve High School Science Students' Motivation to Enroll in a University's Department of Science and Engineering

Yuto Omae, Katsuko T. Nakahira, and Hirotaka Takahashi
Nagaoka University of Technology
Niigata, Japan
e-mail: y_omae@stn.nagaokaut.ac.jp, katsuko@vos.nagaokaut.ac.jp, hirotaka@kjs.nagaokaut.ac.jp

***Abstract*—Many Japanese students pursuing science and engineering majors drop out of the program due to a lack of motivation to learn. One reason is students' declining interest in science or engineering, mainly because their career goals are vague. Helping students make better choices when choosing a university/department by increasing their motivation is an effective way to improve this situation. One approach is to provide stimuli that increase students' interests and provide them with a feeling of satisfaction. Changing the educational environment is one way to provide these stimuli. We assume that students' motivation to study science and technology will increase if we contribute to constructing these environments. In this paper, we propose an educational design including micro-insertion to improve students' motivation.**

Keywords-educational design; perception; motivation; micro-insertion; teaching science.

I. INTRODUCTION

One goal of the Japanese government policy is to realize a "scientific and technological nation," due to the shortage of energy resources. Thus, the Japanese government has conducted many projects for training human resources of technology-focused public educational institutions (e.g., Super Science High School and Science Partnership Program [1][2]). As a result of these projects, the ratio of an enrollment to university in the department of science and engineering was 20% [3]. However, the dropout rate in this department is higher than that in Liberal Arts [4]. The major reasons that students drop out of the program are classified as "positive reasons (e.g., studying abroad, changing interests, aiming at acquisition of qualifications)" and "negative reasons (e.g., loss of motivation, insufficient credits, and student apathy)" [4]. The existences of the dropout students from the department of Science and Engineering become disincentive for the above purpose in Japan. As a result, young people not in education, employment, or training (NEET) and unemployment rates increase [5]. Therefore, it is important to address the negative reasons that students drop out.

To address this issue, it is necessary to increase the motivation of high school students to enroll in university science and engineering. Thus, we focus on the high school science courses taken by the applicants to the department of Science and Engineering at the university level. In such courses, some students are highly motivated to enroll in courses at a university's department of science and engineering, and others are not. The highly motivated students will choose and follow an appropriate path because they have a clear career plan. Those with low motivation will not do so because their career plan is vague. However, if they have many chances to come in contact with positive stimuli that increases their interest and gives them satisfaction, their motivation may increase. An education environment that is constructed using an educational design that has such an effect is one solution to improving students' motivation. Therefore, the instructor should be aware of each student's situation because each student is unique; for example, one student may lack interest, while another student lacks confidence. For this reason, it is important to develop an educational design that includes "adaptive" methods. To improve student's motivation effectively, the teacher must appropriately assess the shortage factors that affect a student's motivation. Thus, it is useful for instructors to assess their students' psychological state (e.g., interest, confidence, and satisfaction).

However, if only one of the teachers assesses the motivation of students, it is jot sufficient to increase their motivation. To effectively increase student's motivation to go to university science and engineering courses, many science teachers have to cooperate and aim to increase the importance factor (such as micro-insertion, mentioned later).

Thus, to increase students' motivation, we propose an educational design by which high school teachers of science-related subjects understand the important factors pertaining to Information Technology (IT) and cooperate with one another. The targeted students are high school students who have low motivation to enroll in university classes in science and engineering.

Section II explains the details of the proposed educational design. Section III presents conclusions and discusses future work.

 ISBN: 978-1-61208-390-2

II. EDUCATIONAL DESIGN

A. Factors Affecting Motivation

A student's motivation for the future is affected by interest, confidence and so on. For example, the student who thinks "I can learn something interesting by enrolling in the course" is likely to be motivated. However, the student who thinks "I cannot succeed in this course" is not likely to be motivated. If some external influence changes students' perception, their motivation also changes. Figure 1 illustrates an example. A new perception of science results from interaction between the current perception and an external influence. In addition, the student's motivation increases due to this new perception.

For example, when students experience uninteresting stimuli, their perception and motivation become negative. When they experience interesting stimuli, their perception and motivation improve (Figure 1). The aim of the present study is to increase students' motivation to enroll in university's department of science and engineering by improving their perception of science. Table I lists factors of perception that affect the motivation to enroll in university science and engineering courses. One theory regarding motivation for the future is the expectancy value model [6]. According to this model, choices for the future are affected by *expectation of success*, *intrinsic value*, and *attainment utility value*. Since *satisfaction* is also motivation for behavior selection [7], we assume that *satisfaction* affects an individual's choices for the future. Therefore, we tentatively adopted *intrinsic value*, *attainment utility value*, *expectation of success*, and *satisfaction* as factors affecting the motivation to enroll in university classes in science and engineering. We refer to these four factors as "*science perception.*" We assume that students' motivation to enroll in science and engineering classes depends on their *science perception*, and that science teachers can change this *perception*. For example, if the teacher for the science subject performs the class to increase *intrinsic value* (e.g., talking to increase curiosity), their *intrinsic value* is increase. This increased *intrinsic value* raises students' motivation to enroll in university science and engineering classes. Thus, it is important for science teachers to improve their students' *perception* of science. We propose the use of *micro-insertion* to accomplish this goal.

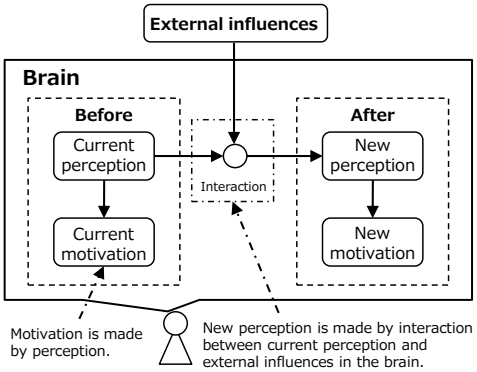


Figure 1. Change of perception and motivation.

TABLE I. FACTORS OF SCIENCE PERCEPTION AND ITS DEFINITIONS

Factor	Definition
Intrinsic value	The enjoyment one gets from engaging in the task or activity
Attainment utility value	The instrumental value of the task or activity for helping to fulfill another short or long-range goal and the link between the task and one's sense of self and identity
Expectation of success	Confidence about one's personal efficacy to master the task
Satisfaction	The degree of feelings such as "I want to do this task" after the task or activity about science in high school.

B. Micro-insertion

Micro-insertion is a teaching method proposed by Michael Davis for engineering ethics [8]. This method involves not only teaching engineering ethics as a major subject but also inserting ethical topics in other classes. The object of our research is not engineering ethics. However, this teaching method such as aiming at achievement purpose of the education with a part of lesson time in many subjects would be applicable to not only engineering ethics but also other purpose on education. Therefore, we include *micro-insertion* in our educational design. Figure 2 (a) depicts typical education, and Figure 2 (b) depicts education using *micro-insertion*. Sample students in Figure 2 (a) and (b) have high enough factors except *attainment utility value*. In other words, the teacher should increase *attainment utility value* for these students. For example, in typical education, the mathematics teacher sets a high value on *attainment utility value*; the physics teacher sets a high value on

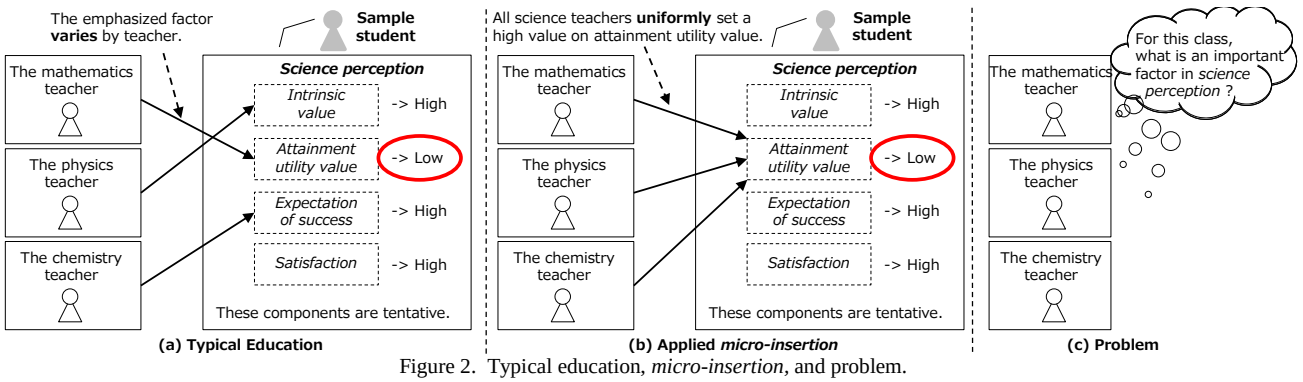


Figure 2. Typical education, *micro-insertion*, and problem.

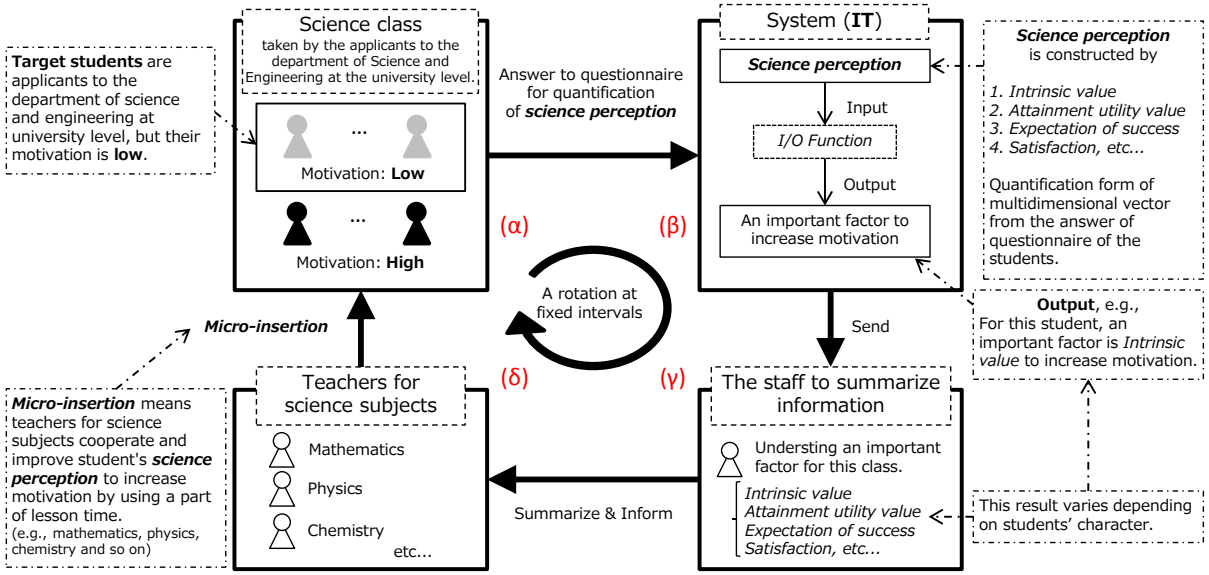


Figure 3. Educational design.

intrinsic value; and the chemistry teacher sets a high value on *expectation of success*. The emphasized factor varies by teacher, as depicted in Fig. 2 (a). Figure 2 (b) presents an example of a teaching method using *micro-insertion*. Unlike typical education, all science teachers uniformly set a high value on *attainment utility value* in their lessons. Because students are provided stimuli uniformly by all science teachers, we assume that the motivation to enroll in a university's department of science and engineering will increase as a result of the use of *micro-insertion* (Figure 2 (b)), unlike the situation with the typical educational approach (Figure 2 (a)).

C. Application of IT

One problem with the application of *micro-insertion* in education is illustrated in Figure 2 (c). Figures 2 (a) and (b) present an example of students whose *attainment utility value* increases. However, in reality, some students' *intrinsic value* or *expectation of success* should increase. To understand an important factor from among *science perception* needed to increase student's motivation, teachers must have expert education experience and a lot of time. This is very difficult. To solve this problem, it is necessary to construct a function that outputs the most effective factor in *intrinsic value*, *attainment utility value*, *expectation of success* and *satisfaction* to increase motivation to go to university science and engineering classes when we input student's *science perception*. In addition, we need to construct a method that summarizes each student results in comparison to the whole class results. Therefore, we must develop a system using machine learning and statistics (IT).

D. Proposed Educational Design

Based on the information above, we propose an educational design using IT and *micro-insertion* (Figure 3).

Figure 3 (α) depicts a high school science class of students who have applied to enroll in university science and engineering classes. This class includes both students with low motivation and those with high motivation. The target of our proposed educational design is those with low motivation. We utilize a questionnaire to quantify the *science perceptions* of those students. The results are presented in Figure 3 (β).

(β) in Figure 3 is a system to output an important factor in *science perception* to increase student's motivation. First, to quantify students' *science perceptions*, the system transforms responses to the questionnaire into a multidimensional vector consisting of *intrinsic value*, *attainment utility value*, *expectation of success*, and *satisfaction*. Next, this vector is input into an input-output (I/O) function. Finally, the I/O function outputs an important factor of *science perception*. This result depends on the students' *science perceptions*. For example, some students' motivation is strongly increased due to increasing *intrinsic value*. However, others' motivation depends on increasing *expectation of success*. This result is sent to (γ) in Figure 3.

The staff of (γ) receives the results from (β). They summarize the results and send messages about them to science teachers (e.g., the target class has many students who need *expectation of success* to increase their motivation to enroll in science and engineering classes). The results vary, depending on students' *science perceptions* in class. This result is sent to (δ) in Figure 3.

In Figure 3, (δ) corresponds to teachers of science-related subjects (e.g., mathematics, physics, and chemistry). These teachers follow an education design that includes *micro-insertion* as depicted in Figure 2 (b), referring to the information from the staff of (γ). For example, if many students need increased *intrinsic value* to increase their

 ISBN: 978-1-61208-390-2

motivation to enroll in university science and engineering classes, the teachers of science-related subjects should give them information that will increase their *intrinsic value* as part of the lessons. Likewise, if many students need increased *expectation of success* to increase their motivation to enroll in university science and engineering classes, teachers of science-related subjects should give them information that will increase their *expectation of success* as part of the lessons. Applicants' motivation to enroll in university science and engineering may increase as a result of using this educational design in various science-related classes rather than in only one subject.

The proposed educational design, as illustrated in the process from (α) to (δ) in Figure 3, increases students' motivation.

III. CONCLUSION AND FUTURE WORK

In this paper, we proposed an educational design that increases students' motivation to enroll in university science and engineering. The target is high school students with low motivation in science classes. To realize it, we will implement the following procedure. First, we need to determine the constituent factors of *science perception* that affect students' motivation to enroll in university science and engineering (In this paper, we use *intrinsic value*, *attainment utility value*, *expectation of success* and *satisfaction* as *science perception*). Second, we need to develop a questionnaire to quantify these factors. Lastly, we need to construct a function to output an important factor to increase students' motivation when we input students' science perception.

After resolving the above issue, we will verify the effect of our proposed educational design. We have conducted a questionnaire survey of high school students (n = 120). We are in the process analyzing the acquired data and will report the results in future work.

ACKNOWLEDGMENT

This work was supported in part by JSPS KAKENHI Grant-in-Aid for Scientific Research (No. 24501146). This work was also supported in part by Nagaoka University of Technology Presidential Research Grant (D).

We would like to thank Takako Mitsui, Yoko Tsuchiya and Rai Shukuin at Yamanashi Eiwa Junior and Senior High School for cooperating on our research.

REFERENCES

[1] Surper Science High School. [Online]. Available from: https://ssh.jst.go.jp/ 2014.11.17

[2] Science Pertnership Program. [Online]. Available from: http://www.jst.go.jp/cpse/spp/ 2014.11.17

[3] Ministry of Education, Culture, Sports, Science and Technology, "Basic Research on School" [Online]. Available from: http://www.mext.go.jp/b_menu/toukei/chousa01/kihon/1267995.htm 2014.11.17

[4] C. Uchida, "Daigaku ni okeru kyu-taigaku ryunen gakusei ni kansuru tyousa (The 32th report of survey about the dropout students in the university)", the report of the 32th meeting of the japanese association for college mental health, 2011.

[5] Ministry of health, labourt and welfare, "NEET no zyoutai ni aru wakamoono no zittai oyobi siensaku ni kansuru tyousa kenkyu (Sruvey about actual state of young fellow for NEET)", [Online]. Available from: http://www.mhlw.go.jp/houdou/2007/06/h0628-1.html

[6] J. S. Eccles, "Where are all the women? Gender differences in participation in physical science and engineering", American Psychological Association, 2005, pp. 199-210.

[7] M. Kitajima and K. Naitou, "Syouhisya koudou no kagaku (Science for consumer behavior)", Tokyo Denki University Press, 2010.

[8] D. Michael, "Integrating ethics into technical courses: *Micro-insertion.*" Science and Engineering Ethics 12.4, 2006, pp. 717-730.

 ISBN: 978-1-61208-390-2

On the Role of Contextual Information in the Organization of the Lexical Space

Flavia De Simone,
Roberta Presta

Scienza Nuova Research Centre
Suor Orsola Benincasa University
Naples, Italy
Email: flavia.desimone
@centroscienzanuova.it,
roberta.presta
@centroscienzanuova.it

Simona Collina

Suor Orsola Benincasa University
Naples, Italy
Email: simona.collina
@unisob.na.it

Robert J. Hartsuiker

University of Ghent
Ghent, Belgium
Email: robert.hartsuiker
@ugent.be

***Abstract*—A hotly debated topic in Psycholinguistics concerns the mental representation of words. The current theories about mental lexicon agree on the idea that contextual information plays a crucial role in the organization of lexical knowledge in mind. This paper presents the results of a study we conducted onto two large scale corpora, an Italian one and a Dutch one, aiming at the evaluation of the power of words context in language learning and processing. To this aim, we leverage an outstanding computational model resembling the basic aspects of the internal language formation process. The experiment outcomes show that, starting from the contextual information, it is possible to gain knowledge of even language-specific characteristics.The results corroborate the language-independence of the model we used. We motivated the representativeness of the model also in the light of the current psychological theories.**

Keywords–Distributed Semantic Representation; Contextual Information; Self-Organizing Map; Semantic Map.

I. INTRODUCTION

There is a general agreement that the degree of semantic similarity between two linguistic expressions depends on the similarity of the linguistic contexts in which they appear. While contextual information analysis represents a valuable quantitative method for semantic analysis and lexical resource induction, from a cognitive perspective, it is also supposed to play a causal role in forming general lexical representations.

In distributed models [1], word meaning is typically represented as a vector in a high dimensional space. Semantically similar words tend to cluster in such a space. The more related they are, the closer they are placed. There are two main approaches to generate semantic distributed models: (i) *feature-based* approach and (ii) *corpus-based* approach. To generate a feature-based model, the first step is to ask human subjects to choose a fixed number of words ("features") to describe the considered target words, these words representing the "context". As one of the main drawbacks, they do not work well with closed class of words (such as, for example, determiners and prepositions) and abstract words. On the other hand, corpus-based models are generated precisely in order to start from large scale corpora of words. The hypothesis is that meaning should be constructed based on the statistical co-occurrences of target words in the corpora. As the power of a computational model depends on the capability to capture the mental property of language, a common issue for both the approaches is the limited number of words they considered in their grammatical features. Specifically, both the approaches cluster nouns/objects and verbs/actions but lack in considering the variability through which a speaker describes reality by means for example of action words like "destruction", so far syntactically a noun but semantically describing an event. This is a very important matter of fact, giving that computational models are often taken as a simulation of the cognitive processes involved in using language [2] [3].

In this study, by using a distributed semantics corpus-based approach, we aim at analyzing the lexical-semantic space of words, including action words, in order to specify how these are represented compared to nouns/objects-verbs/actions dimensions and to better specify how the chosen approach is suitable to model language. We consider a distributed semantics corpus-based approach based on the well-known Contextual Self-Organizing Map (SOM) algorithm [4] and analyze the maps resulting from the processing of two large scale corpora, an Italian corpus and a Dutch corpus.

The paper is structured as follows. Related work needed to place this study is presented in Section II. Section III illustrates the algorithm we have leveraged to produce the semantic maps. In Section IV, we delve into the details of the experimental part of this work and comment the obtained results. Conclusion and future work are finally discussed in Section V.

II. RELATED WORK

According to Jackendoff in [5], a word is a long-term memory trace of phonological, syntactic, and semantic information. Particularly, he suggested that this trace *"lists a small chunk of phonology, a small chunk of syntax, and a small chunk of semantics"*. Over the years, this view of the mental lexicon has been enriched by the idea that contextual information plays a crucial role in the organization of lexical knowledge in mind [6]. This hypothesis is supported by empirical evidence: a series of priming experiments showed that verbs [7] and nouns of events [8] prime agents and objects, suggesting that the mental lexicon encodes event-based relations. As suggested by Elman in [9], the assumption that the meaning of a word is never out-of-context is the insight that underlies computational models which derive words representations from statistical

 ISBN: 978-1-61208-390-2

co-occurrences in large-scale corpora. In order to test the richness of contextual information in deriving lexical-semantic representations of words, a computational model of this sort has been tested. The work in [4] has been considered as a reference guide for the practical methodology we have applied in our study. The cited work realizes the corpus-based analysis of an English and of a Chinese corpus by means of the Contextual SOM algorithm (presented in the following) and illustrates as well the Matlab software package we have exploited in the experiments of this work.

III. Contextual Self-Organizing Maps

A self-organizing map [10] is an artificial neural network capable of unsupervised formation of topology-preserving spatial maps capturing input data characteristics. Input data are typically presented to the map in the form of N-dimensional normalized vectors. Each node of the network is characterized by its own coordinates in the $2-$dimensional grid and by a "weight vector" having the same dimension of the input vectors. The SOM is "trained" in order to let node weights progressively resembling, according to a specified similarity metric, the input data. After a sufficient number of training iterations ("epochs"), the node weights in the SOM will have approximated the distribution of the analyzed data by preserving their distance relationships: similar input will be mapped to neighboring nodes, where the mapping consists in the selection of the node having the most similar weight to the considered input. Consequently, thanks to the reduction of the problem dimensionality from N to 2, the map allows for a visual representation of the input distribution and clusters of similar data can be identified by looking ad the corresponding node regions.

A self-organizing *semantic* map is a self-organized map aimed at representing the semantic space of words on a two-dimensional surface [11]. In order to deal with symbolic input, such as words and their contextual information, an ad-hoc pre-processing phase needs to be addressed. As a result of that phase, a distinct N-dimensional unit-length vector will be assigned to each word and the map will be trained on such dataset. The procedure to build such an input dataset is fully detailed and motivated in [11]. We herein report only the main concepts needed to understand the basic rationale behind the performed corpus elaboration. Each input vector is made by two parts: a symbol part, representing a numerical index uniquely associated with the target word, and an attribute part, named the "average context vector" of the target word. As a preliminary step, to each word is assigned a distinct random D-dimensional vector of unit length. For each target word, it is considered its context, i.e. all the words preceding and succeeding the target word in the corpus, together with their co-occurrence values. Then, two D-dimensional vectors are calculated: (i) the weighted average of the random vectors associated to the predecessors, and (ii) the weighted average of the random vectors associated to the successors. The average context vector of the target word is then built as the sequence of the aforementioned vectors, by obtaining this way a $2D$ dimensional vector.

After the training phase, the map is stimulated with the vectors representing the target words: they are built by concatenating the symbol part associated with each word followed by a null attribute part. The "best matching units" (i.e., the nodes with the most similar weight vectors) are then identified and labeled. This process results in the construction of the graphic semantic map, where it is possible to visually observe "similar" words mapped into clustered areas.

IV. Experiments

To the aim of evaluating the power of the contextual information, we have run two experiments onto two different large-scale corpora, an Italian corpus and a Dutch corpus respectively. We have preprocessed such digital corpora and passed them as input to a Contextual SOM in order to analyze the resulting semantic maps and discuss the represented lexical categories. We have leveraged the Matlab software package documented in [4] to run the Contextual SOM algorithm. In the following, details about the analyzed corpora and the adopted procedure are provided. Finally, we discuss the experiments outcomes.

A. Materials

Two large scale corpora have been analyzed. The first corpus is an Italian corpus, extracted by the CoLFIS database [12] including 194624 word tokens with 7065 unique word types. Such a corpus is made by articles from several Italian journals. The second one is a Dutch corpus, precisely an extract of the SUBTLEX-NL corpus [13], consisting of 1047467 tokens with 33962 types. The Dutch corpus is composed by different movie subtitles. Romance and Germanic languages differ in the quantity of syntactic information carried by single words and in the syntactic structure of phrases.

B. Procedure

The experimental process takes different steps, as described in the following:

1) *pre-processing of the corpus;* to run the algorithm, it has been necessary to pre-process the digital corpus by producing two files: a frequency file, in which word types are listed according to their frequency in the corpus, and a second file, in which each word of the corpus is translated into a numerical index corresponding to the number of the word in the frequency list. In the following, only words having an occurrence greater than 5 in the corpus are considered.
2) *vectorization;* in this phase, to each word is associated a normalized $100-$dimensional random vector, as in the preliminary step of the method described in Section III.
3) *generation of the co-occurrence matrix;* the co-occurrence matrix counts the number of times that word *i* precedes or succeeds word *j* in the corpus. The computation of such matrix is needed in order to build for each word the appropriate average context vector part to be used to train the map.
4) *computation of the input vectors and training of the map;* the input dataset is composed as depicted in Section III. We have used a 50x60 map, by inheriting the same parameters settings adopted in [4] for similar analysis. The network has been trained for 200 epochs.
5) *generation of the semantic map;* we have produced a $300-$word map for the Italian corpus and a

 ISBN: 978-1-61208-390-2

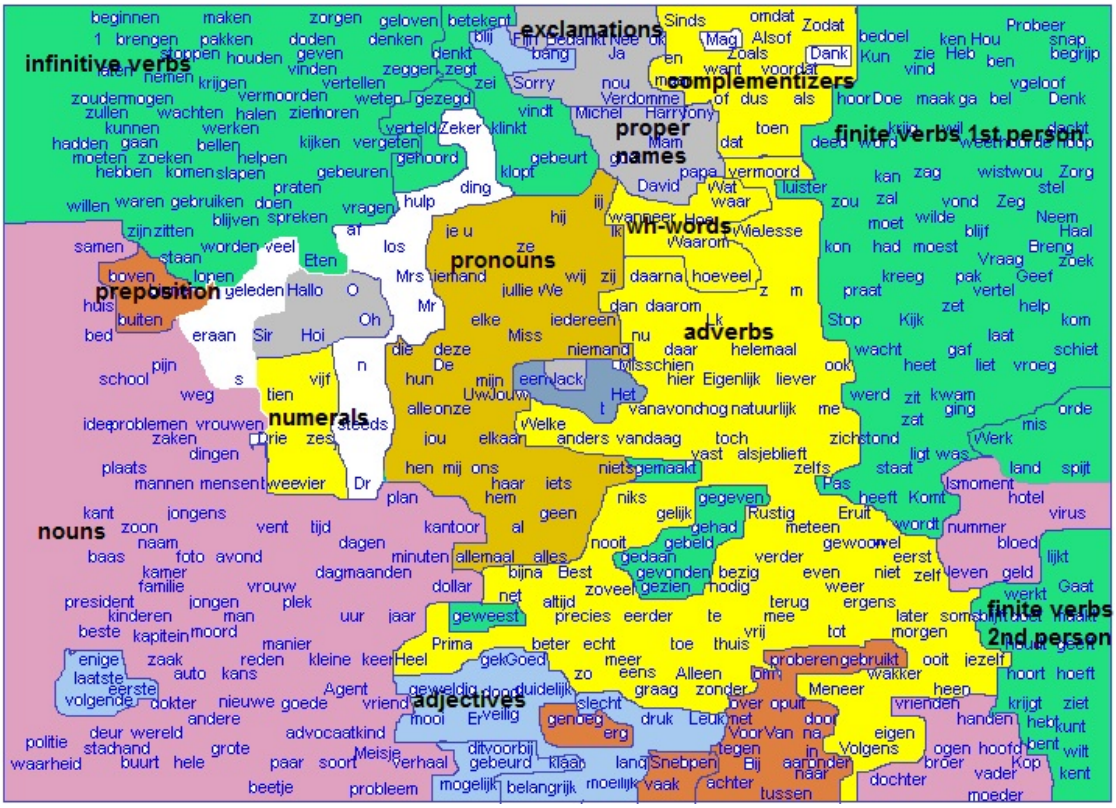


Figure 1. Self-organized map generated from the Dutch corpus

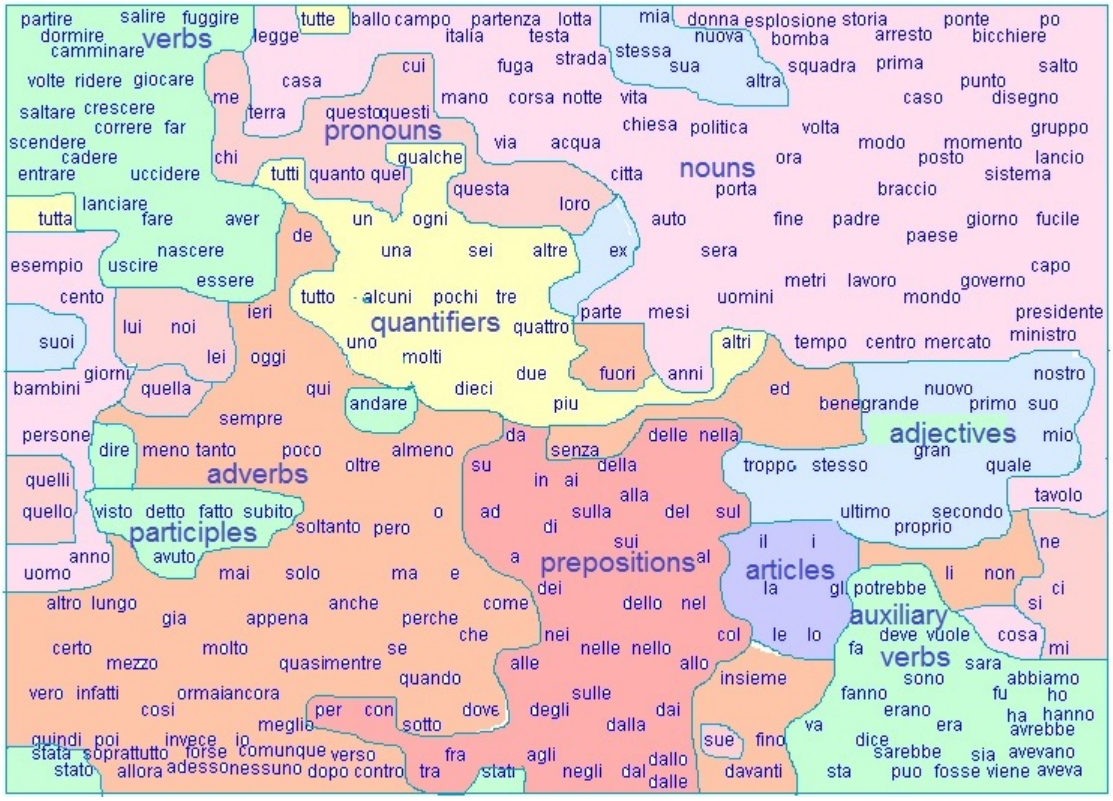


Figure 2. Self-organized map generated from the Italian corpus

500−word map for the Dutch corpus by submitting to the map, respectively, the 300 and 500 most frequent words of the two corpora for the sake of obtaining readable maps.

C. Results

Figure 1 and Figure 2 are the graphic representations of the outcoming semantic maps. We manually draw the boundaries between the semantic clusters in order to make them more clearly visible and to highlight the results. As already mentioned, in the Dutch map just the 500 words with the highest frequencies are presented.

With respect to Figure 1, the first evidence is that the model clusters the major lexical classes, as nouns and verbs, together with closed classes, function words classes as pronouns, preposition and the so-called "wh-words" (why, what, when, where). As suggested in [14], "the closed classes represent a more restricted range of meanings, and the meanings of closed-class words tend to be less detailed and less referential than open-class words." The class of verbs is distributed according to tense (finite vs. infinite), person (1−st and 2−nd), and mood (i.e, all the verbs beginning in capital letters are imperative or interrogative and they are collocated at the margins of the clusters). The class of nouns seems to be organized for gender, with "de-words" in the left part of the map separated from "het-words" in the right part. So far, in the distribution of words in space the lexical-syntactic dimension seems to be more pregnant than the semantic dimension. As to the case of the Italian map, also here we can see that the map clusters the major lexical categories: among the 300 words, it is indeed possible to identify 80 nouns, 24 verbs in the infinitive form, 26 auxiliary verbs, 8 past participles. Unlike in a feature-based model, also closed class as determiners, adverbs, prepositions, and abstract words as "manner" ("modo"), "case" ("caso"), "time" ("tempo") are clustered. Moreover, plural and singular nouns have been clustered separately. All plural nouns, as "men" ("uomini"), "months" ("mesi"), "years" ("anni") are kept together, while all around we can see singular nouns as "center" ("centro"), "work" ("lavoro"), "father" ("padre"). So far, the model is sensitive to semantic and conceptual properties of words. But the network captures also grammatical relations as gender. In the right part of the noun cluster, indeed, there are all masculine words while in the left part there are just feminine words. Also words close in the meaning as "day" ("giorno") and "night" ("notte") are far positioned from each other because they do not share the same grammatical gender. Gender is not only a syntactic property of a word but above all is an arbitrary property. In addition, the map clusters action nouns as "throw" ("lancio"), "jump" ("salto"), "arrest" ("arresto"), "explosion" ("esplosione") with nouns and far from verbs, even if there is a sub-cluster that put them at the margins of network.

V. Conclusion and Future Work

The resulting maps capture the semantic properties of words: semantically similar words are mapped to spatially close positions. More surprisingly, despite the syntactic differences between the two languages, both maps capture syntactic properties as well: grammatical class, mood and tense for verbs, gender for nouns appear as clusters in the visual representations of the networks. Action nouns (like "throw" ("lancio"), "jump" ("salto"), "arrest" ("arresto"), "explosion" ("esplosione")) are clustered with nouns and far from verbs, even if there is a sub-cluster that put them at the margins of network.

These data corroborate the idea that words are not only vectors of semantic information, but also syntactically rich entities, in line with psycholinguistic evidences. The results obtained add a piece of evidence in the long debate on the lexical organization of words and they support a grammatical class distinction that is independent from semantic [15].

In closing, words seem to carry more information than suggested in [5], by this way posing questions about how this further knowledge is stored in mind. The findings of the experimental campaign herein presented fit the view recently expressed in [16] and [9], which challenged the traditional idea of the mental lexicon as a dictionary: *"Rather than putting word knowledge into a passive storage ..., words might be thought of in the same way that one thinks of other kinds of sensory stimuli: they act directly on mental states ..., it is in the precise nature of their causal effects that the specific properties of words phonological, syntactic, semantic, pragmatic, and so forth are revealed"*.

Further data analysis will be necessary to test the potentiality of the algorithm in the simulation of real categorization processes and to discard the hypothesis that the lexical structure in terms of the order of words in the phrases is the only responsible for the organization of the lexical categories in the map. To do so, we will consider the opportunity to use corpora from other languages, like Hebrew for example, where the phrase structure is more flexible and the meaning of a phrase is independent from words order.

Acknowledgment

This research has been performed with support from the EU ARTEMIS JU project HoliDes (http://www.holides.eu/) SP−8, GA No.: 332933. Any contents herein reflect only the authors' views. The ARTEMIS JU is not liable for any use that may be made of the information contained herein. Finally, the authors would like to special thank Professor Marc Brysbaert for his precious support.

References

[1] A. Lenci, "Distributional Semantics in Linguistic and Cognitive Research," Italian journal of linguistics, vol. 20, no. 1, 2008, pp. 1–31.

[2] P. Tabossi, S. Collina, F. Pizzioli, A. Caporali, and A. Basso, "Speaking of Actions: the Case of CM," Cognitive Neuropsychology, vol. 27(2), 2010, pp. 152–180.

[3] S. Collina, P. Marangolo, and P. Tabossi, "The Role of Argument Structure in the Production of Nouns and Verbs ," Neuropsychologia, vol. 39, no. 11, 2001, pp. 1125 – 1137.

[4] X. Zhao, P. Li, and T. Kohonen, "Contextual Self-Organizing Map: Software for Constructing Semantic Representations," Behavior research methods, vol. 43, no. 1, 2011, pp. 77–88.

[5] R. Jackendoff, Foundations of Language: Brain, Meaning, Grammar, Evolution. Oxford University Press, USA, 2002.

[6] J. J. Van Berkum, C. M. Brown, P. Zwitserlood, V. Kooijman, and P. Hagoort, "Anticipating Upcoming Words in Discourse: Evidence from ERPs and Reading Times," Journal of Experimental Psychology: Learning, Memory, and Cognition, vol. 31, no. 3, 2005, p. 443.

[7] T. R. Ferretti, K. McRae, and A. Hatherell, "Integrating Verbs, Situation Schemas, and Thematic Role Concepts," in Journal of Memory and Language, 2001, pp. 516–547.

 ISBN: 978-1-61208-390-2

[8] M. Hare, M. Jones, C. Thomson, S. Kelly, and K. McRae, "Activating Event Knowledge," Cognition, vol. 111, no. 2, 2009, pp. 151–167.

[9] J. L. Elman, "Lexical Knowledge without a Lexicon," in Mental Lexicon, 2011, pp. 1–33.

[10] T. Kohonen, Ed., Self-organizing Maps. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 1997.

[11] H. Ritter and T. Kohonen, "Self-organizing Semantic Maps," Biological Cybernetics, vol. 61, no. 4, 1989, pp. 241–254.

[12] P. M. Bertinetto, C. Burani, A. Laudanna, L. Marconi, D. Ratti, C. Rolando, and A. M. Thornton, "Corpus e Lessico di Frequenza dell'Italiano Scritto (CoLFIS)," 2005, http://linguistica.sns.it/CoLFIS/Home.htm, [accessed: 2015-02-12].

[13] E. Keuleers, M. Brysbaert, and B. New, "SUBTLEX-NL: A New Measure for Dutch Word Frequency Based on Film Subtitles," Behavior research methods, vol. 42, no. 3, 2010, pp. 643–650.

[14] M. L. Murphy, Lexical Meaning. Cambridge University Press, 2010.

[15] F. De Simone and S. Collina, "The Picture-Word Interference Paradigm: Grammatical Class Effects in Lexical Production," Journal of Psycholinguistic Research, 2015, submitted.

[16] J. L. Elman, "An Alternative View of the Mental Lexicon," in Trends in Cognitive Sciences, 2004, pp. 301–306.

 ISBN: 978-1-61208-390-2

Implementing Relational-Algebraic Operators for Improving Cognitive Abilities in Networks of Neural Cliques

Ala Aboudib, Vincent Gripon and Baptiste Tessiau
Télécom Bretagne - Electronics Department
Brest cedex 3, France
Emails: ala.aboudib@telecom-bretagne.eu, vincent.gripon@telecom-bretagne.eu, baptiste.tessiau@ens-rennes.fr

***Abstract*—Associative memories are devices capable of retrieving previously stored messages from parts of their content. They are used in a variety of applications including CPU caches, routers, intrusion detection systems, etc. They are also considered a good model for human memory, motivating the use of neural-based techniques. When it comes to cognition, it is important to provide such devices with the ability to perform complex requests, such as union, intersection, difference, projection and selection. In this paper, we extend a recently introduced associative memory model to perform relational algebra operations. We introduce new algorithms and discuss their performance which provides an insight on how the brain performs some high-level information processing tasks.**

Keywords–Cognitive modeling; artificial neural networks; relational algebra; associative memory; simulated annealing

I. INTRODUCTION

Associative memories are special types of memories that are capable of high-speed content-based mapping between input queries and outputs. This type of storage differs from classical index-addressable memories in that no explicit address is needed to search for stored data. In order to efficiently provide this associative functionality to memories, different methods of content structuring are required. Artificial neural networks (ANNs) are known to be such an adapted medium to implementing associative memorization. Their design is inspired from neo-cortical neural mechanisms in mammalian brains, believed to be knowledge-associative biological neural networks [1].

Many ANN models, differing in topology and functionality, were proposed to act as associative memories. Famous examples include the Perceptron [2], self-organizing maps [3], Hopfield networks [4] and Boltzmann machines [5]. A new model was proposed recently by Gripon and Berrou in [6] and generalized by Aliabadi et al. in [7]. This model relies on sparse coded patterns stored as cliques and is based on Hebian learning [8]. This sparse coding approach resembles the ones introduced by willshaw [9] and Palm [10], with an added explicit mapping between stored messages and their representation in the network.

Typically, an associative memory is able to retrieve a previously stored piece of information given partial content, a sort of erasure-retrieval property. However, human ability to handle more complex queries suggests that their representation of information should be able to perform other operations. Relational algebra gives a formal framework to introduce complex operations on tuples of stored messages, including union, intersection, difference, projection and selection. In this paper, we aim at extending the model introduced in [7] to process these operations.

Most of these operations are common as far as human cognition is concerned. For example, selection aims at retrieving the list of all messages that match a given probe (e.g., listing all city names you know that start with the letter 'b'). The union operation can be used for merging date while intersection and difference operation are useful for comparing contents of several memories or memory regions. Our main motivation is to show that existing neural-network-based architectures for associative memories are easily adapted in order to handle these complex queries. There have been several works addressing the relational problem and its biological plausibility such as [11] and [12], which were more focused on inference and relational learning.

The rest of this paper is organized as follows: in Section II, we introduce the associative memory model extended in this paper. Sections III, IV, V and VI describe how to perform resp. union, intersection, difference and projection using these models. In Section VII, we explain how to handle the more complex selection operator. For this operator, we introduce a novel algorithm using simmulated annealing. Simulation results are provided in Section VIII when independently identically uniformly distributed messages are considered. Section IX is a conclusion.

II. THE MEMORY MODEL AND RELATIONAL ALGEBRA

A. The associative memory model

First, we introduce the associative memory model proposed by Gripon and Berrou in [6] and then extended in [7]. Let us consider the finite alphabet $\mathcal{A}$ made of integers between 1 and ℓ. we define a blank character $\perp \notin \mathcal{A}$ and $\overline{\mathcal{A}} = \mathcal{A} \cup \{\perp\}$. We are interested in storing words made of χ characters over $\overline{\mathcal{A}}$. We call such a word a *message* $m = m_1 m_2 \dots m_\chi$. Blank characters represent the absence of a character at a given position, such that depending on their number, messages can be regarded as sparse vectors.

An associative memory is a device capable of storing messages and then retrieving them given partial knowledge about some of their nonblank characters. To implement this device, the authors of [6] propose to use a symmetric, binary, χ-partite neural network composed of $n = \chi \cdot \ell$ vertices that we shall refer to as *units*. The authors of [6] explain

 ISBN: 978-1-61208-390-2

that units should be considered analogous to cortical micro-columns believed to be the computational building-blocks of the cerebral neo-cortex [13] [14]. This network can be split into χ *clusters*, each containing the same number ℓ of units.

We denote by $[n]$ the set of integers beween 1 and n, let us then index each cluster of the network by an integer in $[\chi]$. We also index units of a given cluster by integers in $[\ell]$. As a result, each unit in the network is uniquely addressed giving a couple (i, j) where i is the index of a cluster and j the index of the unit within this cluster.

We define a function f that maps each message m into a set of couples (i, j) as follows:

$$f : m \mapsto \qquad \{(i,j) \big| i \in [\chi], j \in [\ell] \text{ and } m_i = j\} \tag{1}$$

where m_i refers to the i-th character of m.

The network topology is entirely captured by an adjacency matrix W of size $\chi \cdot \ell$ such that:

$$w_{(i,j)(i',j')} = \begin{cases} 1 & \text{if } (i,j) \text{ and } (i',j') \text{ are connected} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

To store a message $m = m_1 m_2 \dots m_\chi$ in such a network, all unit pairs in $f(m)$ are connected pairwise according to 2 forming a clique in the underlying graph. It is worth noticing that cliques corresponding to different messages could sometimes overlap and share connections. Given a set of messages $\mathcal{M}$, we denote by $W(\mathcal{M})$ the adjacency matrix obtained after storage of all messages in $\mathcal{M}$.

A subset of $f(m)$ is called a *partial input* associated with the message m. The task of an associative memory is then: given a partial input of $m \in \mathcal{M}$, retrieve m using $W(\mathcal{M})$.

To retrieve a message from a partial input, an iterative algorithm is performed. Retrieval algorithms and techniques have been discussed in detail in [15].

It has been shown in [7] that performance of this structure as an associative memory mainly depends on a density parameter defined as the ratio of the number of connections in the network to the total number of possible connections. Under the hypothesis that messages contain exactly c non-blank characters uniformly distributed over $\overline{\mathcal{A}}$, density can be approximated by the following equation:

$$d = 1 - \left(1 - \frac{c(c-1)}{\chi(\chi-1)l^2}\right)^M \tag{3}$$

In this paper, we aim at extending the functionality of these associative memories, which we shall call Clustered Clique Networks (CCNs), to cover more general problems defined in relational algebra.

B. Connections to relational algebra

In relational algebra, operators are defined on *relations* (sets of tuples). A set of attributes is associated with each relation. Then, each tuple is defined as a set of instances of these attributes. A CCN in this respect can be viewed as a relation. Each cluster represents an attribute and units within each cluster are instances of that attribute (attribute values). A clique connecting units is equivalent to a tuple. In the following sections, we are going to propose algorithms and methods for implementing some relational operators on relations defined in the form of CCNs.

III. UNION

Defined in terms of the set theory, the union of a collection of sets $S^1, S^2, ..., S^n$ is a set $S^\cup$ containing all distinct elements in this collection. It can be described as follows:

$$S^\cup = S^1 \cup S^2 \cup ... \cup S^k \tag{4}$$

$$= \{x | x \in S^1 \ \vee \ x \in S^2 \ \vee \ ... \ \vee \ x \in S^k\} \tag{5}$$

where $\cup$ is the set union operator and $\vee$ is the Boolean OR function.

In the context of memory storage, union is used to combine the contents of several memories or memory partitions into a single one while avoiding the redundancy resulting from the same data-word being stored multiple times. An example of this is merging the contents of two folders on a computer. This task is straightforward when using a classical indexed memory since messages do not overlap.

Suppose that we have two CCNs with the same dimensions $W(\mathcal{M}_1)$ and $W(\mathcal{M}_2)$ that we wish to merge in a single network $W^\cup$ of the same dimensions. We define this operation as follows:

$$w^\cup_{(i,j)(i',j')} = \begin{cases} 1 \text{ if } \begin{matrix} w(\mathcal{M}_1)_{(i,j)(i',j')} = 1 \vee \\ w(\mathcal{M}_2)_{(i,j)(i',j')} = 1 \end{matrix} \\ 0 \text{ otherwise} \end{cases} \tag{6}$$

Following intuitively from this definition is the fact that upon applying the union operation, information is conserved in the new network. That is, if a clique exists in either $W(\mathcal{M}_1)$ or $W(\mathcal{M}_2)$, it would also exists in $W^\cup$. Hence we can rewrite $W^\cup$ as $W^\cup(\mathcal{M})$ with $\mathcal{M} = \mathcal{M}_1 \cup \mathcal{M}_2$. The problem here is that combining memories in this fashion can cause a significant growth in the density of $W^\cup(\mathcal{M})$, leading possibly to dramatically low performance in terms of retrieval error rates of stored messages. More formally, if d^1 is the density of $W(\mathcal{M}_1)$ and d^2 is the density of $W(\mathcal{M}_2)$, then the density of $W^\cup(\mathcal{M})$ denoted $d^\cup$ is given by:

$$d^\cup = 1 - (1 - d^1)(1 - d^2) \tag{7}$$

and thus $d^\cup$ is greater than both d^1 and d^2. This means that while retrieval error rates may be optimal for $W(\mathcal{M}_1)$ and $W(\mathcal{M}_2)$, the network resulting from their union might suffer from some degeneration in performance because of the increased density. Therefore, union should be done only if the resulting network performance is acceptable. The relationship between retrieval error rates and density are presented in details in [7] and [15].

According to (6) all possible connections in the networks should be tested during the union operation. So, given $\frac{\chi(\chi-1)\ell^2}{2}$ possible connections [7] in each network, the average-case complexity of such process is $\Theta(\chi^2\ell^2)$.

The phenomenon of degenerated memorization efficiency caused by the increased density is not uncommon in the brain. For example, learning and recalling a new word in a foreign language is typically not a difficult task. However, trying to learn a dozen of new words at the same time might turn out to be much more challenging comprising many memorization errors and confusions and even mixing syllables of different words. New words can be better learned by training and experience, which is partially due to the association of these words with other memories. We suggest that this process, in

 ISBN: 978-1-61208-390-2

terms of CCNs is equivalent to adding more clusters to the network such that more units can be added to existing cliques, which lowers its density and increases performance.

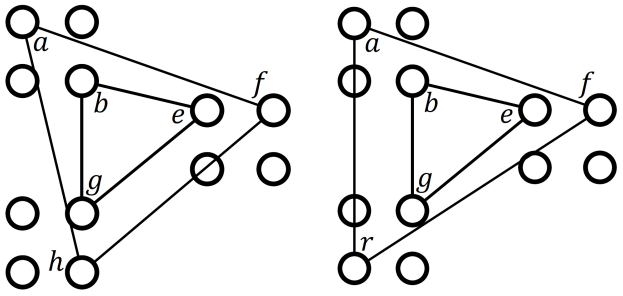
IV. INTERSECTION

The intersection among several sets $S^1, S^2, ..., S^k$ is a set $S^\cap$ containing only those elements that $S^1, S^2, ..., S^k$ have in common:

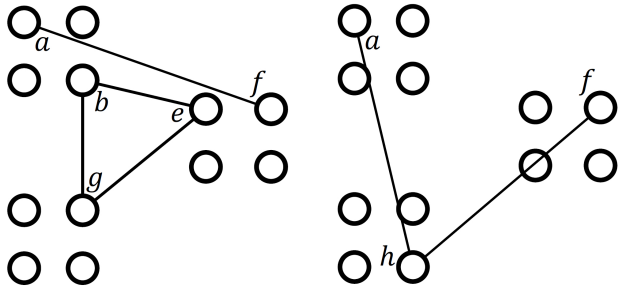
$$S^\cap = S^1 \cap S^2 \cap ... \cap S^k \quad (8)$$

$$= \{x | x \in S^1 \ \wedge \ x \in S^2 \ \wedge \ ... \ \wedge \ x \in S^k\} \quad (9)$$

where $\cap$ is the set intersection operator and $\wedge$ is the Boolean AND function.



(a) Network $W(\mathcal{M}_1)$ (b) Network $W(\mathcal{M}_2)$



(c) Network $W^\cap$ (d) Network $W^\setminus$

Figure 1. Intersection and Difference between two CCNs. $W^\cap$ shown in 1c is the network resulting from the intersection of $W(\mathcal{M}_1)$ and $W(\mathcal{M}_2)$. $W^\setminus$ shown in 1d is the difference of $W(\mathcal{M}_1)$ from $W(\mathcal{M}_2)$.

When applied to relations, intersection serves in extracting common tuples between two or more tables having the same number and types of attributes. This is done easily when such a database is stored in an indexed memory. One way to implement intersection between two CCNs is by keeping only those connections that happen to exist in the exact same place in both networks. So, given two CCNs $W(\mathcal{M}_1)$ and $W(\mathcal{M}_2)$ of the same type and dimensions, we can define their intersection $W^\cap$ as follows:

$$w^\cap_{(i,j)(i',j')} = \begin{cases} 1 \text{ if } & w(\mathcal{M}_1)_{(i,j)(i',j')} = 1 \wedge \\ & w(\mathcal{M}_2)_{(i,j)(i',j')} = 1 \\ 0 \text{ otherwise} \end{cases} \quad (10)$$

The average-case complexity of this operation is given by $\Theta(\chi^2\ell^2)$ for the same reason as in the union. The density $d^\cap$ of $W^\cap$ as a function of d^1 and d^2 (the densities of $W(\mathcal{M}_1)$ and $W(\mathcal{M}_2)$ respectively) is given by:

$$d^\cap = d^1.d^2 \quad (11)$$

We notice from (11) that $d^\cap$ is always lower than d^1 and d^2 given that densities have their values in the interval [0,1]. Thus, we guarantee that no density explosion occurs as in union.

Some problems might still occur when performing this operation, as depicted in Figure 1 where four simple identical networks are considered each with a total number of units $n = 12$ grouped in $\chi = 3$ clusters with a message size of $c = \chi$. Network $W(\mathcal{M}_1)$ shown in Figure 1a contains two cliques afh and beg and network $W(\mathcal{M}_2)$ in Figure 1b also contains two cliques afr and beg. We notice that only the clique beg is common between these two networks. By applying the intersection operation as in (10) we obtain $W^\cap$ shown in Figure 1c, which contains the common clique beg as expected but also contains the edge af, which is an undesirable result, because af does not represent a complete message (tuple). We shall call af a *residual* edge. As a consequence, $W^\cap$ could possibly contain more connections than a an ideal intersection network $W^\cap(\mathcal{M})$ with $\mathcal{M} = \mathcal{M}_1 \cap \mathcal{M}_2$. This increase in density due to residual edges is expected to deteriorate performance [7].

V. DIFFERENCE

The difference of two sets S^1 and S^2, which can also be called that relative complement of S^2 with respect to S^1, is a set $S^\setminus$ that contains only those elements of S^1 that are not in S^2:

$$S^\setminus = S^1 \setminus S^2 = \{x | x \in S^1 \text{ and } x \notin S^2\} \quad (12)$$

where $\setminus$ is the set difference operator which is not commutative so that $S^1 \setminus S^2 \neq S^2 \setminus S^1$.

So, the difference between two database tables is the set of tuples in the first one that do not exist in the other, which is also an easy-to-implement operation in classical memory systems. A simple method of implementing difference between two CNNs $W(\mathcal{M}_1)$ and $W(\mathcal{M}_2)$ is by instantiating a new memory $W^\setminus$ containing only connections in $W(\mathcal{M}_1)$ that do not exist in $W(\mathcal{M}_2)$. That is:

$$w^\setminus_{(i,j)(i',j')} = \begin{cases} 1 \text{ if } & w(\mathcal{M}_1)_{(i,j)(i',j')} = 1 \wedge \\ & w(\mathcal{M}_2)_{(i,j)(i',j')} = 0 \\ 0 \text{ otherwise} \end{cases} \quad (13)$$

The average-case complexity of this operation is also given by $\Theta(\chi^2\ell^2)$. Intuitively, the density $d^\setminus$ of the new network $W^\setminus$ is always less than or equal to the density d^1 of $W(\mathcal{M}_1)$. So, if d^1 is well controlled, we would have no problems with the performance of $W^\setminus$. The density $d^\setminus$ is given by the following relationship:

$$d^\setminus = d^1.(1 - d^2) \quad (14)$$

As in the case of intersection, the difference operation defined in (13) processes data down on the level of individual connections not on the level of cliques. This causes the problem depicted in Figure 1d. In this figure, $W^\setminus$ is the network resulting from applying $W(\mathcal{M}_1) \setminus W(\mathcal{M}_2)$. Ideally, we wish that $W^\setminus = W^\setminus(\mathcal{M})$ with $\mathcal{M} = \mathcal{M}_1 \setminus \mathcal{M}_2$,i.e., a network that contains only the clique afh. However, according to (13), we only get two edges ah and fh because af is a common edge between $W(\mathcal{M}_1)$ and $W(\mathcal{M}_2)$ and thus eliminated by the difference operation. This represents a loss of information since the network $W^\setminus$ no more stores the message

 ISBN: 978-1-61208-390-2

corresponding to the clique afh. We shall call this undesirable effect *erosion*. Actually, no method is yet available for getting an ideal intersection or an ideal difference between CCNs. We consider the methods proposed in this paper as approximations to the real operations.

VI. PROJECTION

In relational algebra, projection is defined as a unary operator Π applied to a tuple R in order to produce a new tuple R^{Π} consisting of k attributes $\{r_1, r_2, ..., r_k\}$, which is a subset of the attributes originally contained in R. This can be written as follows:

$$R^{\Pi} = \Pi_{r_1, r_2, ..., r_k}(R) \tag{15}$$

A network W^{Π} is said to be a projection of a network W on a given set of attributes, if it contains only a subset of the attributes of W ,i.e., W^{Π} contains only a subset of the clusters of W.

Clearly, this operation is very easy to implement in CNNs with a constant time complexity. Moreover, the density d^{Π} of the resulting network is equal to the density of the original network given that connections are uniformly distributed within the latter network:

$$d^{\Pi} = d \tag{16}$$

VII. SELECTION

A. Problem definition

In relational algebra, selection or restriction is a unary operator applied to a relation R^1 and returns another relation R^2. The latter relation contains all tubles in R^1 whose attribute values satisfy a propositional formula φ. This can be transcribed as follows:

$$R^2 = \sigma_{\varphi}(R^1) \tag{17}$$

We argue in this paper that a mechanism similar to selection might be used by the brain for thinking and memorization. A typical such request would be, for instance, to name all scientific authors one knows whose names start with an 'a'. We aim at using the model proposed in [7] as a substrate for this selection process.

The selection algorithm we shall present here runs continuously, giving multiple (possibly redundant) answers one after another. This appears to us being behaviorally similar to what humans could produce facing a similar query. The process that makes us avoid redundancy is called short-term/working memory. Once its buffer is full or overcrowded, repetitions of the same word/name can occur.

The requirements that our selection algorithm is meant to meet are the following:

1) Determine the sub-graph G of potentially interesting units.
2) In sub-graph G, find all cliques of size c.

Finding a maximum clique in a graph (or equivalently) a minimum cover is a know NP-complete problem. Many algorithms and heuristics were proposed to give approximate solutions to this problem in medium-sized graphs. Examples of these algorithms are [16] and [17] that make use of the simulated annealing principle introduced in [18] and [19], which is a probabilistic meta-heuristic method for locating the global optimum of a given function. Another known method widely used in applications such as computational chemistry is the BronKerbosch algorithm [20], which can efficiently find maximal cliques in an undirected graph.

We propose to use an adaptation of the simulated annealing algorithm proposed by Geng et al. in [16]. We also use their same objective function to evaluate our solutions.

B. The proposed selection algorithm

Suppose we have a CCN denoted by W and a partial input message m containing $q \leqslant c$ known nonblank attributes. The selection operator consists in finding all stored cliques made of a set of units containing $f(m)$. In order to perform this search efficiently, it is sufficient to restrict the search to the subgraph made of only the units connected to units in $f(m)$. We shall call this subgraph G_m. We denote its adjacency matrix by A_m.

For simplicity of representation, we refer to each unit of G_m by an integer index k or s where $k, s \in \{0, 1, ..., n' - 1\}$, n' being the number of units in G_m. Our objective now is to find all the cliques in G_m that have a size (number of units) of $c' = c - q$.

We consider the following optimization problem: We define ρ as a permutation of units in G_m (ρ is an array of size n' containing all unit indexes of G_m as its elements). For a given ordering of elements in ρ, we consider the first c' elements of ρ as indexes of units in G_m acting as a potential solution (a clique). So, by permuting ρ's contents we can get a different candidate solution. The objective function used to evaluate these solutions is given by:

$$F(G_m, \rho) = \sum_{k=0}^{c'-2} \sum_{s=k+1}^{c'-1} (1 - a_{\rho[k], \rho[s]}) \tag{18}$$

where $a_{\rho[k], \rho[s]}$ is an element of the adjacency matrix W. As $F(G, \rho) = 0$ when a clique is found, the goal is to find permutations that minimize this function.

The algorithm is applied to G_m as follows:

Step 1: Parameter initialization.
Initial temperature T_1, end temperature T_2, current temperature $t = T_1$ and cooling coefficient α. Set the initial permutation as $\rho[k] = k, k \in \{0, 1, ..., n' - 1\}$. Random setting of permutation is also possible.

Step 2: Compute $F(G_m, \rho)$ and terminate if it evaluates to zero.

Step 3: Randomly choose two integer indexes u and w of ρ such that $u \in \{0, 1, ..., c' - 1\}$ and $w \in \{c', c' + 1, ..., n' - 1\}$.
Condition 1:
If $\rho[w]$ has more or the same number of connections with $\{\rho[0], \rho[1], ..., \rho[c' - 1]\}$ than $\rho[u]$ has, then $\rho[w]$ and $\rho[u]$ are swapped and thus a new permutation ρ' is obtained.
Condition 2:
If $\rho[u]$ has more connections with $\{\rho[0], \rho[1], ..., \rho[c' - 1]\}$ than $\rho[w]$ has, then the index w is rejected and we go back to step 3. but if w has already been rejected for more than $8n'$ times consecutively as defined in [16], then

$\rho[w]$ and $\rho[u]$ are swapped and a new permutation ρ' is obtained.

Step 4: Compute $F(G_m, \rho')$ and terminate if it evaluates to zero.

Step 5: Compute $\Delta F = F(G_m, \rho') - F(G_m, \rho)$. If $\Delta F \leqslant 0$, then accept the new permutation ρ' by setting $\rho = \rho'$. Otherwise accept ρ' with probability $p = e^{\frac{-\Delta F}{t}}$.

Step 6: Update current temperature as $t = \alpha t$. If $t < T_2$ terminate the algorithm; otherwise, go back to step 3.

A single run of this selection algorithm is meant to find one clique ,i.e., one answer. So, in order to output more answers this algorithm should be repeated many times.

VIII. RESULTS

A. Selection

In order to test the proposed selection algorithm, we used a CCN with $n = 3240$ units grouped in $\chi = 15$ clusters. 15000 randomly generated messages of $c = 10$ characters were stored giving a network density d of about 0.13. We used the same initial and end temperatures as in [16] for the simulated annealing algorithm; 100 and 0.001, respectively. We set the cooling coefficient to 0.996. To construct a selection query message, one already stored message is randomly selected of which 9 characters were erased (set to $\perp$) giving an input query message with only one known non-blank character.

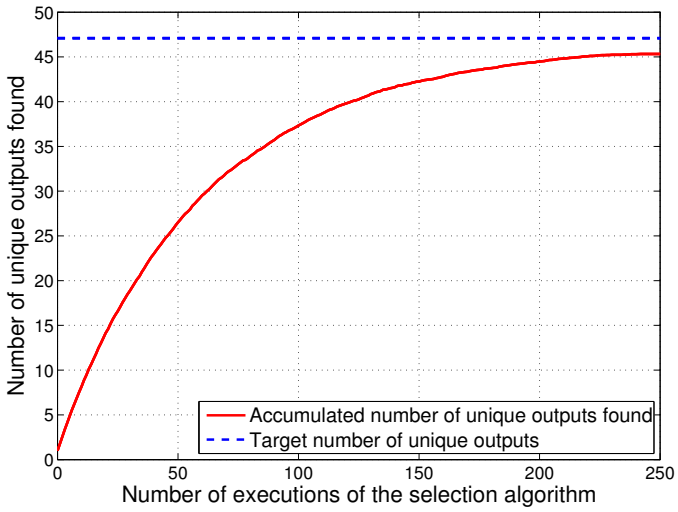


Figure 2. Average accumulated number of unique outputs as a function of the number of iterations in a network of $n = 3240$, $\chi = 15$, $c = 10$, 15000 stored messages. The input message used has only one non-blank character.

As described in Section VII, the algorithm we propose might give redundant outputs when executed several times. Figure 2 shows how fast unique outputs were found as a function of the number of executions of the selection algorithm. Curves in the figure are averaged over 100 identical experiments.

An interesting property of the resulting curve is that most unique answers are obtained during earlier executions of the algorithm. So, referring back to Figure 2, about 45 unique answers out of 47 possible ones are found by the 250^{th} execution, 40 answers by the 125^{th} execution and about 30 answers by the 60^{th}. In other words, about 89% of answers were obtained when only 50% of total executions were achieved and 67% of answers after 24% of total executions. This is a natural result for a selection by replacement experiment where all answers have an equal chance of being chosen at each execution.

We suggest that such result bears some qualitative resemblance to the way human beings memorize lists of mental objects where it is common that the last few items turn out to be more difficult and time consuming to recall because of the distraction caused by redundant answers coming to mind and other phenomena.

B. Intersection and difference

A comparison among average complexities of some operators when applied to CCNs and two other known data structures (ordered lists and binary search trees) storing sparse messages of the form $m = m_1 m_2 \dots m_\chi$ is provided in Table I. An interesting observation is the fact that the order of complexity of operators using a CCN is close to that of a binary search tree given that setting $\ell \gg \chi$ is preferable in practice for a higher network capacity [7]. However, the complexity of union, intersection and difference of ordered lists is lower by a factor of χ than that of CCNs while the complexity of insertion is ℓ^2/χ higher.

TABLE I. COMPLEXITY OF SOME RELATIONAL OPERATORS IN SEVERAL TYPES OF DATA STRUCTURES.

	CCN	ordered list	binary search tree
Insertion(Storing)	$\Theta(\chi^2)$	$\Theta(\chi\ell^2)$	$\Theta(\chi \log(\ell))$
Union	$\Theta(\chi^2\ell^2)$	$\Theta(\chi\ell^2)$	$\Theta(\chi\ell^2 \log(\ell))$
Intersection	$\Theta(\chi^2\ell^2)$	$\Theta(\chi\ell^2)$	$\Theta(\chi\ell^2 \log(\ell))$
Difference	$\Theta(\chi^2\ell^2)$	$\Theta(\chi\ell^2)$	$\Theta(\chi\ell^2 \log(\ell))$

IX. CONCLUSION AND FUTURE WORK

In this paper, we have introduced some methods for applying certain algebraic-relational operators on a new class of neural-network-based associative memories we call CCNs. We argued that the process of recalling a list of items (which can also be mapped to more general memorization tasks) in the brain can be behaviorally assimilated to the selection operation known to relational algebra. We proposed an algorithm for implementing this process using the principle of simulated annealing. Then, we showed that the results we got have some resemblance to what might be obtained by a human subject in terms of redundancy.

We have also demonstrated that CCNs can be used as classic data-structures by approximating operators such as union, intersection, difference and projection. We saw that the implementation of union and its related density explosion problem raised the question as to how the brain organizes information with high correlation or high density. Two possible mechanisms the brain might be using are forgetting rarely used "data" (by the decay of synaptic weights) and tagging pieces of correlated data with different contextual information. Similar mechanisms might be integrated in CCNs by allowing connections to have real values with a decay parameter and by providing contextual tagging in the form of CCNs existing on

 ISBN: 978-1-61208-390-2

a separated level of a hierarchy of networks. Another possible solution is to design networks with dynamic sizes to prevent exceeding a maximum allowed density.

ACKNOWLEDGMENT

This work was supported by the European Research Council under the European Union's Seventh Framework Program (FP7/2007-2013) / ERC grant agreement n 290901.

REFERENCES

[1] J. R. Anderson and G. H. Bower, Human associative memory. Psychology press, 2013.

[2] F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain." Psychological review, vol. 65, no. 6, 1958, p. 386.

[3] T. Kohonen, "Self-organized formation of topologically correct feature maps," Biological cybernetics, vol. 43, no. 1, 1982, pp. 59–69.

[4] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," Proceedings of the national academy of sciences, vol. 79, no. 8, 1982, pp. 2554–2558.

[5] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, "A learning algorithm for boltzmann machines," Cognitive science, vol. 9, no. 1, 1985, pp. 147–169.

[6] V. Gripon and C. Berrou, "Sparse neural networks with large learning diversity," Neural Networks, IEEE Transactions on, vol. 22, no. 7, 2011, pp. 1087–1096.

[7] B. K. Aliabadi, C. Berrou, V. Gripon, and J. Xiaoran, "Storing sparse messages in networks of neural cliques." IEEE transactions on neural networks and learning systems, vol. 25, no. 5, 2014, pp. 980–989.

[8] D. O. Hebb, The Organization of Behavior. John Wiley, 1949.

[9] D. J. Willshaw, O. P. Buneman, and H. C. Longuet-Higgins, "Non-holographic associative memory." Nature, 1969.

[10] G. Palm, "On associative memory," Biological Cybernetics, vol. 36, no. 1, 1980, pp. 19–31.

[11] H. Blockeel and W. Uwents, "Using neural networks for relational learning," ICML-2004 Workshop on Statistical Relational Learning and its Connection to Other Fields, 2004, pp. 23–28.

[12] J. E. Hummel and K. J. Holyoak, "A symbolic-connectionist theory of relational inference and generalization." Psychological review, vol. 110, no. 2, 2003, p. 220.

[13] E. G. Jones, "Microcolumns in the cerebral cortex," Proceedings of the National Academy of Sciences, vol. 97, no. 10, 2000, pp. 5019–5021.

[14] V. B. Mountcastle, "The columnar organization of the neocortex." Brain, vol. 120, no. 4, 1997, pp. 701–722.

[15] A. Aboudib, V. Gripon, and X. Jiang, "A study of retrieval algorithms of sparse messages in networks of neural cliques," COGNITIVE 2014, The Sixth International Conference on Advanced Cognitive Technologies and Applications, 2014, pp. 140–146.

[16] X. Geng, J. Xu, J. Xiao, and L. Pan, "A simple simulated annealing algorithm for the maximum clique problem," Information Sciences, vol. 177, no. 22, 2007, pp. 5064–5071.

[17] X. Xu and J. Ma, "An efficient simulated annealing algorithm for the minimum vertex cover problem," Neurocomputing, vol. 69, no. 7, 2006, pp. 913–916.

[18] S. Brooks and B. Morgan, "Optimization using simulated annealing," The Statistician, 1995, pp. 241–257.

[19] V. Černỳ, "Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm," Journal of optimization theory and applications, vol. 45, no. 1, 1985, pp. 41–51.

[20] C. Bron and J. Kerbosch, "Algorithm 457: finding all cliques of an undirected graph," Communications of the ACM, vol. 16, no. 9, 1973, pp. 575–577.

 ISBN: 978-1-61208-390-2

In Pursuit of Natural Logics for Ontology-Structured Knowledge Bases

Jørgen Fischer Nilsson

Department of Mathematics and Computer Science
Technical University of Denmark (DTU)
Email: jfni@dtu.dk

***Abstract*—We argue for adopting a form of natural logic for ontology-structured knowledge bases with complex sentences. This serves to ease reading of knowledge base for domain experts and to make reasoning and querying and path-finding more comprehensible. We explain natural logic as a development from traditional logic, pointing to essential differences to description logic. We conclude with a knowledge base set-up with an embedding into clausal logic, offering also a graph view of the sentences.**

***Index Terms*—Knowledge representation and reasoning; Ontologies; Natural language Interface.**

I. INTRODUCTION

We address ontology-structured knowledge bases (KBs), that is KBs which encompass ontological classification structures as well as more general logical sentences. We outline a KB set-up which supports sentences in a regimented (controlled) fragment of natural language. This choice is motivated by our wish to make knowledge bases readable for domain experts. Moreover, this approach offers "generative ontologies" where linguistic terms generate new concept nodes in the ontology in addition to the given classes. In the devised meta logic set-up, the natural logic KB sentences are embedded in a clausal logic taking care of the inference and querying.

Our approach to ontological engineering is described further in recent papers [1][2][3][4]. We focus on knowledge bases within the life-sciences, which abound with complex textual descriptions and elaborate classification structures.

A. State of the Art

Contemporary approaches to knowledge based systems aim at accommodating more complex information than admitted in traditional relational databases. The two competing prominent approaches are the rule-based representations in the form of logical clauses and various dialects of description logic. These logical representations are described and compared e.g. in Grosof et al. [5].

As an alternative to these logics for KBs we apply so-called natural logic for the considered KBs including formal ontologies. The applied natural logic possesses a logical semantics and is supported by reasoning rules applied directly to the natural logic form. Sentences in the natural logic resemble natural language, so that KBs can be read and understood by domain experts. Moreover, as a novelty, the set-up simultaneously provides a graph representation of the natural logic KB content as extension of the common ontological partial order classification diagrams. The supporting graph representation facilitates pathfinding in a KB. This functionality enables computation of shortest paths in the KB graph between user-stated concepts and entire phrases.

B. The Structure of the Paper

The structure of this paper is as follows: In Section II, we take as departure traditional syllogistic logic. Then, in Section III we review *en passant* essentials of description logic as a tool for setting up formal ontologies. In Section IV, we turn to our main subject of natural logics, followed up, in Section V, by introduction of the natural logic fragment we propose for ontological knowledge bases. For the implementation set-up for the natural logic dialect we consider the logic of definite clauses in Section VI, which is used for embedding of the natural logic knowledge base in the devised KB systems design in Section VII. Finally, Section VIII concludes the paper.

II. TRADITIONAL SYLLOGISTIC LOGIC AND ONTOLOGIES

Let us begin recalling the Aristotelian natural logic syllogistic sentence forms [6] known from the square of opposition, see figure 1.

every C isa D	no C isa D
some C isa D	some C isa not D

Fig. 1 From the square of opposition.

Contemporary formal ontologies apply basically the class inclusion relation isa corresponding to the sentence form every C isa D, which forms a partial order by way of reflexivity, transitivity and antisymmetry. Often, the partial order, rendered as a Hasse diagram (graph), simplifies to a classification tree with the universal top class at the root.

Although there are (unspecified) extension sets behind the classes, usually there is no requirement that the ontology has to form a distributive lattice, let alone a lattice by presence of supremum and infimum classes [7]. This is because the intensional comprehension of classes makes in particular many would-be union classes ontologically irrelevant [8].

As for the other three forms above in the square of opposition, in ontological engineering they are often expressed by introducing appropriate classes in the ontology, together with the default assumption that classes are disjoint if they have no common subclass and neither is a subclass of the other. Recall that traditional logic comes with existential import,

 ISBN: 978-1-61208-390-2

meaning that there is no notion of empty classes. Individual concepts may formally be conceived of as singleton classes in the ontological set-up.

III. Description Logic

Description logic (DL), the foremost contemporary knowledge representation logic, is a fragment of predicate logic. DL has become pivotal as logical basis for the semantic web research endeavors. The various description logic dialects share a variable-free algebraic form of expressions, with the general requirement that the logic is decidable with respect to desired functionalities, as well as tractable. In Grosof et al. [5], DL is compared to and aligned with the rule forms in definite clause logic. The tractability requirement implies that intended operations can be performed in polynomial time measured in terms of the size of the description logic specification. The various dialects of description logic differ by the admitted operators and the ensuing worst case complexity.

A. Description Logic and Ontologies

The ontological class inclusion relationship "C isa D" in description logic becomes

$$C \sqsubseteq D \tag{1}$$

In DL, there are no default rules such as the above mentioned existential import. Accordingly, disjointness of two classes is expressed, for instance by

$$C \sqcap D \equiv \bot \tag{2}$$

where $\bot$ is the predefined empty concept (class).

Classes C and D in DL generalize to various concept expression forms including set union, $\sqcup$, and intersection $\sqcap$. As such the ontological constitution in DL provides distributive lattices. Furthermore, even Boolean lattices are achieved by complement formation. In this way, DL offers class generativity in formal ontologies.

B. Concept Modifiers

Description logic offers means of forming sub-classes (called concepts in DL) notably by means of a binary algebraic operator $\exists R.C$, where the first argument, R, is binary relation (a property in DL terminology), and the second one, C, is a concept expression. For instance, the concept of "*cells that produce insulin*", being a sub-concept of "*cell*", becomes

$$cell \sqcap \exists produce.insulin$$

From the point of view of ontological constitutions, the recursive syntactical form of the $\exists$ construct provides potentially unbounded generativity into ever more specialized concepts in the ontology.

Turning from concepts to entire assertions, the sample sentence "*cells that produce insulin reside in the pancreas*" in DL may become

$$cell \sqcap \exists produce.insulin \sqsubseteq \exists residein.pancreas$$

which seems hard to interpret for most domain experts, not the least because of the awkward 'subject-property' copula form, corresponding to "*cells and produce insulin are [something that] reside in the pancreas*".

IV. Natural Logics

Natural logics are formal logics taking form of "regimented" fragments of natural language with accompanying inference rules for reasoning directly with the natural logic [9][10][11][12]. Quoting from the discussion of natural logic in [13]: "*The idea of the universality of logic is based on the conviction that [...] there are certain invariant features of human reasoning, carried out in any natural language whatsoever, that allow the formulation of universal logical laws, applicable to any language.*" In our setup, rather than translating the natural language forms into, say, DL, we conduct reasoning at the natural logic level, unlike Azevedo et al. [14].

A. Class Relationships versus Property Ascriptions

Natural logics may be viewed as a development of traditional syllogistic logic continued via medieval logicians, e.g., John Buridan, see Klima [13][15], and via 19th century logicians, notably Peirce and De Morgan, see e.g. Sánchez Valencia [12]. A key point in this development is the abandoning of strict copula forms taking form of a subject and a predicate as in traditional syllogistic logic, in favour of logical sentences admitting a main verb expressing a binary relationship. In a more conceptual or ontological view, what is at stake here is acceptance of binary point relationships between classes rather than property ascription to classes. The latter "monadistic" view attributed to Leibniz, cf. [16], remains in DL.

As an example, in the property ascription view, informally the sentence *betacell produces insulin* is coined into the somewhat awkward *betacell isa (producer-of insulin)* possibly accompanied by the reciprocal *insulin isa (produced-by betacell)*.

V. A Natural Logic For Knowledge Bases

Let us consider the natural logic in our [3][4] with sentences of the syntactic form

$$Q_1\ C\ R\ Q_2\ D \tag{3}$$

– where Q_1 and Q_2 are either of the determiners (quantifiers) every and some, and
– where C and D are nominal phrases, and
– where R is a transitive verb.

In the simplest case, C and D are common nouns representing classes. These common nouns may next be adorned with modifiers in the form of linguistic relative clauses and adjectives. Modifiers are here assumed to act restrictively, unlike parenthetical relative clauses [13].

As it appears, the form (3) comprises four quantifier combinations, dubbed $\forall\forall, \forall\exists, \exists\forall, \exists\exists$ in [17]. From the point of view of ontological knowledge bases, the by far most common quantifier constellation among the 4 combinations is the $\forall\exists$ option

$$\text{every}\ C\ R\ \text{some}\ D \tag{4}$$

Example: every betacell produces some insulin

 ISBN: 978-1-61208-390-2

– or, in short, using common default conventions as in natural language for this quantifier case:

betacell produce insulin

The existential quantification over substances such as insulin we ontologically understand as ranging over the conceivable amounts of the substance. These amounts constitute the imaginable individuals in a substance class.

One observes that the corresponding passive voice sentence [every] insulin isproducedby [some] betacell, with the reverse relation, is not logically equivalent [17]. However, the weaker some insulin isproducedby every betacell is entailed, adopting existential import throughout.

As an aside, one may notice that the considered form (4) fits perfectly with the partonomic forms in [18] with examples

[every] pancreas haspart [some] betacell

and

[every] betacell ispartof [some] pancreas.

A. Inclusion in the Natural Logic and in Description logic

In the considered natural logic dialect, the class inclusion comes about with the above $\forall\exists$ form with the relation being equality. Thus

every betacell is-equal-to some cell

expresses the class inclusion

betacell isa cell.

This follows from a predicate logical explication of "every C $equals$ some D" as $\forall x(C(x) \rightarrow \exists y(x = y \wedge D(y))$, which is logically equivalent to $\forall x(C(x) \rightarrow D(x))$. We retain the distinguished short form isa relationship in our natural logic, since this relationship prevails in ontological knowledge bases.

The key natural logic sentence "every C R some D" in DL would become the somewhat awkward copula (subject predicate) form

$$C \sqsubseteq \exists R.D \tag{5}$$

cf. Section III.B.

B. Compound Concept Terms

As mentioned, in the devised natural logic [2][3][4] class expressions may contain modifiers with restrictive relative clauses, as in the example

cell that produce insulin residein pancreas

contrast the DL formulation in Section III.B.

VI. Definite Clause Logic: Rule Language

We now turn to definite clause rules as an additional component in our natural logic KB set-up. The building blocks of logical clauses are atomic formulas $p(t_1, ..., t_m)$, where p is an m-argument predicate and the t_i are logical terms. In the present context, these terms are confined to variables and constants representing individuals. Recall that the more general form of clauses applied in logic programming and artificial intelligence admits functional terms, consisting of a function symbol followed by term arguments.

As such, clausal logic specifies relationships between individuals, unlike the focus on relationships between concepts in syllogistic logic and description logic. This makes clausal logic *prima facie* unfit for ontology-structured knowledge bases dealing with relationship between classes.

A logical clause is a disjunction of atomic formulas or their denials, where all variables present (if any) are implicitly universally quantified. A definite clause is conveniently written and understood as an implication clause

$$p_0(t_{01}, ..., t_{0m_0}) \leftarrow \bigwedge_i^n p_i(t_{i1}, ..., t_{im_i}) \tag{6}$$

where the reverse implication arrow can be read as "if". The case of $n = 0$ yields an atomic formula, (called a fact if it is variable-free). Definite clauses where the terms are either variables or constants are known as DATALOG. A logical computation is initiated with an atomic formula as hypothesis to be confirmed or disconfirmed as logically entailed by the given clauses. The DATALOG logic enjoys properties of decidability (proved by propositionalization, i.e., reduction to propositional logic) and tractability, cf. [5].

Definite clauses only express assertive (positive) propositions. However, denials may be provided implicitly by the adoption of the closed world assumption, implying that the denial of a fact is taken to hold if the fact does not follow from the given clauses, a principle known as negation-by-failure (to prove). The DATALOG logic enjoys properties of decidability (proved by propositionalization, i.e., reduction to propositional logic) and tractability, cf. [5].

A. Concepts Reified as Individuals

Definite clauses at the outset express relations between individuals as in

$$hormone(X) \leftarrow insulin(X)$$

However, by encoding of concepts as individuals definite clauses can emulate class-class relationships as in

$$isa(insulin, hormone)$$

supported by the clauses

$$isa^*(C, D) \leftarrow isa(C, D)$$
$$isa^*(C, D) \leftarrow isa(C, X) \wedge isa^*(X, D)$$
$$isa^*(C, C)$$

where isa now represents the immediate (direct) subclass relationship, and the predicate name isa^* its reflexive transitive closure computed in DATALOG. This encoding of classes suggests as a next crucial step embedding of the entire natural logic in clauses with supporting inference directly in the natural logic.

VII. An Embedded Natural Logic

We now wish to embed the natural logic in DATALOG clauses acting as a metalogic for the natural logic. This

calls for decomposition of natural logic sentences into atomic components which can be handled in DATALOG. We devise a decomposition that enables reconstruction of the natural logic compound sentences (modulo paraphrasing) [3][4].

As an instructive example, consider again the sentence

cell that produce insulin residein pancreas

This sentence in our system becomes decomposed into the following atomic ground (that is, variable-free) facts

$$isa(cell'that'produce'insulin, cell)$$
$$fact(definition, cell'that'produce'insulin, produce, insulin)$$
$$fact(observ, cell'that'produce'insulin, residein, pancreas)$$

where $cell'that'produce'insulin$ is formally an auxiliary individual constant in DATALOG, and simultaneously a fresh concept node label in the ontological graph view.

The two latter DATALOG facts, as it appears, comprise an epistemic mode tag. These modes affect the inference engine: The outlet definitions (including isa) of a concept node effectively act as an if-and-only-if definition, unlike the observational contributions. Further modes may be introduced, in order to distinguish normative, observational and hypothetical contributions.

This decomposition principle supports the graph view of ontologies with the subclass relationships forming the skeleton ontology, as it were.

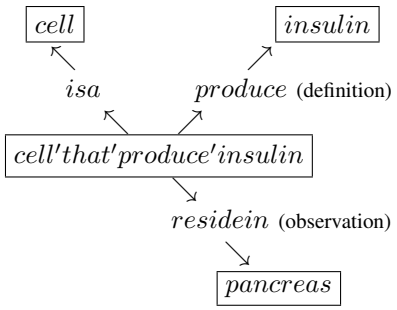


Fig. 2. Graph view of sample sentence.

Figure 2 shows the graph conception of the considered sentence, where the decomposition into the above three ground atomic facts appears as labeled, directed edges being outlet from the concept node $cell'that'produce'insulin$.

The epistemic distinctions ensure that the natural logic sentence is recoverable. They also ensure that relevant subsumption relationships can be computed and added to the KB [4]: Suppose that it is stated that insulin is a hormone in the KB ontology. Then the concept, say, $cell'that'produce'hormone$ is likely to occur also. In this case, the subsumption algorithm then is to compute and record the inclusion relationship $isa(cell'that'produce'insulin, cell'that'produce'hormone)$, concomitant with an additional arc in the graph. On the other hand, we refrain from pre-computing and storing those inclusion isa relationships holding solely by virtue of transitivity, since the entire transitive closure relation would shortcut paths which might preferably be retained in pathway computations.

The compound natural logic sentences in the KB in general give rise to auxiliary nodes in the graph. And the graph contributions from the sentences form a single ontological KB graph with unique node representation of concepts across the sentence contributions. The original natural logic sentences can be reconstructed relying on the edge modes. Ideally, synonymic phrases such as pancreatic cell and cell that residein pancreas would be mapped into one concept node in the KB. One should also keep in mind that all the edges are here further assumed $\forall\exists$-quantified.

A. Intensional Querying and Pathfinding

The embedded knowledge base may now be queried deductively via the clause language, appealing to appropriate inference rules expressed as clauses.

Given class names, c, are introduced by

$$class(c)$$

The concepts (simple or complex) may be queried, say, with

$$\leftarrow isa^*(X, c)$$

giving for variable X all concept terms below c,

– or more restrictively with

$$\leftarrow class(X) \wedge isa^*(X, c)$$

giving all subordinate class names.

The key inference rules in natural logic are the so-called monotonicity rules [9], which admits restriction of the grammatical subject concept to sub-concepts (recognized as inheritance), and, conversely, generalization of the grammatical object concept for the $\forall\exists$ forms considered here [4]:

$$fact(M, Csub, R, Dsup) \leftarrow isa^*(Csub, C) \wedge fact(M, C, R, D) \wedge isa^*(D, Dsup)$$

It follows logically, for instance, given cell that produce insulin residein pancreas and pancreas isa endocrinegland that cell that produce insulin residein endocrinegland.

In [2][3] we discuss pathway inference computations in natural logic KBs in the context of bio-models. This functionality aims at finding shortest paths in the KB graph between two stated concepts appealing to graph search algorithms. Pathway computations may formally be understood as application of a logical comprehension principle for composition of relations [1].

B. Class Disjointness in the Natural Logic

As it stands, the present natural logic does not provide negation, unlike the classical negation available in DL. However, some form of negation is achievable by appeal to negation by non-provability in the rule logic, as known from logic programming and relational database querying.

In our set-up, two classes are considered disjoint unless one class is a sub-class of the other one, or that they have

 ISBN: 978-1-61208-390-2

a common sub-class. We find this natural in ontological engineering, which often leans towards hierarchical classifications. Recall that classes here are assumed non-empty according to the principle of existential import. Then, overlap of two classes (concepts) can be ascertained with

$$overlap(C, D) \leftarrow isa^*(X, C) \wedge isa^*(X, D)$$

where the variable X ranging over concept terms (including those stemming from the decomposition of sentences) may be considered existentially quantified to the right of the reverse implication.

Conversely the disjointness of two classes is verified with

$$disjoint(C, D) \leftarrow \text{NOT } overlap(C, D)$$

appealing to negation by non-provability, NOT, with the closed world assumption, conforming with use of negation in database query languages. From the point of view of ontology development use of the non-monotonic negation by non-provability implies that extension with new overlapping classes to the knowledge base may cancel out present class disjointness.

VIII. Concluding Summary

We have advocated the adoption of forms of natural logic for ontology-structured knowledge bases in a set-up with embedding into definite clauses. This embedding of the natural logic sentences facilitates useful functionalities such as intensional reasoning and querying and pathway finding in large knowledge bases.

We conduct evaluation with a small scale prototype written in the logic programming language PROLOG (supporting DATALOG as a sublanguage) on life-science sample KBs in [2]. The prototype decomposes the natural logic sentences into the shown fact/graph KB representation. The devised decomposition of the natural logic sentences with inference rules in DATALOG invites as a next development step large-scale implementation on relational data base platforms with the decomposed KB sentences represented as tuples.

References

[1] T. Andreasen, H. Bulskov, J. Fischer Nilsson, P. Anker Jensen, and T. Lassen, "Conceptual pathway Querying of Natural Knowledge Bases from Text Bases", 10th International Conference on Flexible Query Answering Systems, FQAS 2013, H. L. Larsen et al (ed.), LNAI 8132, Springer, 2013, pp. 1-12.

[2] T. Andreasen, H. Bulskov, J. Fischer Nilsson, and P. Anker Jensen, "Computing Pathways in Bio-Models Derived from Bio-Science Text Sources", International Conference on Bioinformatics and Biomedical Engineering, F. Ortuño and I.Rojas (Eds.), IWBBIO 2014, April 7-9, Granada (Spain), 2014, pp. 217-226.

[3] T. Andreasen, H. Bulskov, J. Fischer Nilsson, and P. Anker Jensen, "Computing Conceptual Pathways in Bio-Medical Text Models", Foundations of Intelligent Systems - 19th International Symposium, ISMIS 2014, Roskilde, Denmark, June 28-30, LNAI 8502, Springer, 2014, pp. 264-273.

[4] T. Andreasen and J. Fischer Nilsson, "A Case for Embedded Natural Logic for Ontological Knowledge Bases", 6th International Conference on Knowledge Engineering and Ontology Development, KEOD2014, Rome, September 21st-24th, 2014, paper no. 71.

[5] B. N. G. Grosof, I. Horrocks, R. Volz, and S. Decker, "Description Logic Programs: Combining Logic Programs with Description Logic", in Proceedings of the 12th international conference on World Wide Web (WWW '03). ACM, New York, NY, USA, 2003, pp. 48-57.

[6] J. L. Ackrill (Ed.), Aristotle's Categories and De Interpretatione, Oxford at the Clarendon Press, 1963.

[7] J. Fischer Nilsson, "Ontological Constitutions for Classes and Properties", 14th Int. Conference on Conceptual Structures, ICCS 2006, Lecture Notes in Artificial Intelligence, LNAI 4068, Springer, 2006, pp. 37-53.

[8] D. M. Armstrong, A Theory of Universals: Volume 2: Universals and Scientific Realism, Cambridge Universty Press, 1978, 1990.

[9] J. van Benthem, Essays in Logical Semantics, Studies in Linguistics and Philosophy, Vol. 29, D. Reidel Publishing Company, 1986.

[10] J. van Benthem, "Natural Logic, Past And Future", Workshop on Natural Logic, Proof Theory, and Computational Semantics 2011, CSLI Stanford, URL: http://web.stanford.edu/~icard/logic&language/2011.NatLog.pdf [accessed: 2015-01-20].

[11] R. Muskens, "Towards Logics that Model Natural Reasoning, Program Description Research Program in Natural Logic", URL: http://lyrawww.uvt.nl/~rmuskens/natural/ NLprogram.pdf [accessed: 2015-01-20].

[12] V. Sánchez Valencia, "The Algebra of Logic", in D. M. Gabbay and J. Woods (Eds.), Handbook of the History of Logic, Vol. 3 The Rise of Modern Logic: From Leibniz to Frege, Elsevier, 2004.

[13] G. Klima, "Natural Logic, Medieval Logic and Formal Semantics", URL: http://filozofiaiszemle.net/wp-content/uploads/2012/06/Gyula-Klima-Natural-Logic-Medieval-Logic-and-Formal-Semantic.pdf [accessed: 2015-01-20].

[14] R. R. de Azevedo, F. Freitas, R. Rocha, J. A. Menezes, and L. F. A. Pereira, "Generating Description Logic $\mathcal{ALC}$ from Text in Natural Language", in Proceedings of Foundations of Intelligent Systems - 21th International Symposium, ISMIS 2014, Roskilde, Denmark, June 25-27, LNAI 8502, Springer, 2014, pp. 305-314.

[15] G. Klima, John Buridan, Oxford University Press, 2009.

[16] J. Fischer Nilsson, "On Reducing Relationships to Property Ascriptions", Information Modelling and Knowledge Bases XX, Volume 190 Frontiers in Artificial Intelligence and Applications, (Eds.) Y. Kiyoki *et al.*, IOS Press, Hardcover ISBN: 978-1-58603-957-8, 2009, pp. 245-252.

[17] J. Fischer Nilsson "Diagrammatic Reasoning with Classes and Relationships", A. Moktefi and S.-J. Shin (Eds.), Visual Reasoning with Diagrams, Studies in Universal Logic, Birkhäuser, Springer, pp. 83-100, 2013.

[18] B. Smith and C. Rosse, "The Role of Foundational Relations in the Alignment of Biomedical Ontologies", MEDINFO 2004, M Fieschi et al. (Eds.), Amsterdam IOS press, pp.444-448, 2004.

 ISBN: 978-1-61208-390-2

Modeling Situation Awareness: The Impact of Ecological Interface Design on Driver's Response Times

Thomas Friedrichs, Andreas Lüdtke
R&D Division Transportation – Human Centered Design
OFFIS – Institute for Information Technology
{friedrichs, luedtke}@offis.de

Abstract— **Endsley's Situation Awareness (SA) theory and a variety of SA-measurement methods like SAGAT and SPAM aim to explain how humans make errors and assess the SA of operators in dynamic workspaces. However, in order to evaluate the impact of future assistance systems on the SA of operators at design time, predictions about operator performance are needed. In this work, existing SA measurement methods are used to construct a cognitive model which predicts driver reaction times on the basis of SA to road and system events. Ecological Interface Design variants will be used as a test case to show how information presentation influences driver performance.**

Keywords-situation awareness; cogntivie systems; evaluation; ecological interface design; response time; dynamic systems.

I. INTRODUCTION

Truck platooning is defined as a series of trucks that drive with close distances and automatic longitudinal control. In our case, all trucks are equipped with a cooperative cruise control system which communicates with other trucks in order to cooperatively control the distances and the speed of each truck. Braking maneuvers are executed automatically. All drivers maintain lateral control all the time. The driver of the lead truck uses a conventional cruise control system (with optional sensor based braking assistance) and observes the driving scene. The lead truck driver is also responsible for emergency braking actions. The benefits of driving in a platoon include reduced fuel consumption, better use of the infrastructure and improved safety. Bergenheim et al. provide an overview of platooning systems [1].

However, drivers in a platoon are not in a fully automated setting where no manual actions are needed. They are required to constantly steer the truck. Furthermore, they have to regain full control over the vehicle very quickly if necessary. This can happen if the system reaches its functional limits or a road hazard requires immediate intervention of the driver. Drivers also have to remember when platooning maneuvers like splitting, merging or expanding will happen and need to receive sufficient support with these tasks [2]. Another aspect is that if the system makes actions (e.g., adapts the speed automatically), the driver should not be surprised [2], which is an issue known as "Automation Surprises" [3].

Figure 1. Close-following scene in a platoon.

It is therefore important that the driver maintains a sufficient level of Situation Awareness (SA) which is defined as "*[…] the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future*" [4]. SA incorporates three levels: Level 1 is the perception of information, Level 2 is information integration and Level 3 is the projection of the future status. In platooning, the short inter-vehicle distances lead to problems with the visual perception (SA Level 1) of the environment because of the back of the trailer directly in front of the driver (see Figure **1**). Thus, drivers miss crucial visual information. Another aspect is that the lateral control is with the driver, while longitudinal control is a system function. Drivers might get bored and uninformed about the driving situation because they are by default not required to drive fully manual (SA Level 2 and 3). Both factors can contribute to human-out-of-the-loop problems because the SA of the driver is impaired. If an operator gets removed from the control loop, the responses get slower and breakdowns or malfunctions might occur [2][5]. Short response times are a crucial factor in driving, because even fractions of a second can make the difference between an accident and avoidance. For example, a truck which moves with a velocity of 80 km/h covers a distance of 22,22 m in one second.

It is assumed that if the driver is in the loop and therefore has a sufficient SA, response times are minimal. Response time is here defined as the time interval from where a certain

 ISBN: 978-1-61208-390-2

stimulus is perceived to where a possible action can be executed. It does not include the time from onset until perception, time for motor movement or task completion. It is therefore a "cognitive" response time.

It becomes evident that drivers in a platoon need to receive support in order to maintain a good SA because of the occlusion they are not able to do this themselves. An approach is to offer platooning support in the form of an information system which serves as a "third eye". It supports the drivers so that they can maintain a sufficient level of SA. Such a system will be referred to as a Platooning Support System (PSS). It consists of a HMI that displays information about the system state and the driving context. This information can include:

- Sensor readings
  - Distance readings
  - Vehicle velocity
- Environmental information
  - Weather information
  - Topographic information
  - Road status
- Platooning / navigation information
  - Merge / split maneuvers
  - Accordion maneuvers
  - Route information
  - System status, future actions

It is not sufficient to only display this information; drivers have to understand the variables and the interplay between them in an easy way. A promising way to do this, is to display continuous information about the changes and linkages of relevant information from the environment [6]. This approach is grounded in the margins of the Ecological Interface Design framework, which proved to increase operator knowledge in complex and dynamic environments [7]. For example in electric cars, often the flow of energy is visualized. If the car runs in electric power mode, energy flows from the battery to the engine and when the car brakes, energy flows from the brakes back to the battery. This information presentation offers a good way of informing the driver about a variety of parameters and the status of a complex system in an easy to understand and efficient way.

Continuous information presentation can give drivers the ability to track the changes in an evolving situation, which in turn can lead to a better understanding of the dynamics of the situation. This approach uses functional information of the situation, which is relevant to the driver's goals. In terms of SA, continuous information supports on Level 2 and 3 because the comprehension (Level 2) and projection (Level 3) are supported. This is important because if drivers can anticipate what either the system does in the future or how the road status changes, it is assumed that the response times of the driver to these events decrease. Results from other domains show that such a support can increase SA even in unanticipated situations [8]. PSS therefore should use the continuous information approach for information relevant for safe driving. Relevant variables can include relative distances to the surrounding cars, relative speeds, system status and changes, future system actions and so on.

In contrast to continuous information support, there are situations where immediate actions and warnings are needed. For example, if there is a pressure decrease in one of the trucks tires, the driver needs to get informed immediately. Here, continuous information would be inappropriate. This is because tire pressure rarely gets into a dangerous state and therefore the effort to keep track of continuous information would be too high. This approach is referred to as event-based information presentation. Here, drivers are only informed about a certain status change by a warning sound, message or other indication in the cockpit when an immediate action is necessary. For a close-following scenario in the platooning context, certain information can be visualized using this approach, for example blind spot warnings or changes of the speed limit. A balanced combination of continuous and event-based information presentation is a promising approach to support drivers' SA in close-following and at the same time it reduces the complexity of HMIs.

However, when designing such a system it is not clear if information is better presented in a continuous or event-based way. The traditional way of testing such a system would involve experts, focus groups and most importantly, human testers. The latter are not only the most valuable source of feedback, from a legal perspective, human testing is a requirement for the homologation of a new product like a PSS. The evaluation effort for a PSS is very high, because it is much effort to plan, conduct and analyze tests with human testers. Moreover, dynamic situations are very complex and a variety of scenarios have to be covered, what makes this approach even more complex.

Therefore, the objective of this work is to create a cognitive model of driver SA which predicts response times of a cognitive agent under the influence of visual interface design variants. For this, a traffic simulation is used, which includes a platooning scenario. Experiments with real drivers will be performed to calculate the model fit. Two different HMI designs, which will be developed in the scope of the COMPANION project [9], will serve as test cases. One design will show functional information in a continuous way, the other design will use event-based warnings.

In Section 2, the current state of the art of cognitive modeling in dynamic contexts is described. Section 3 presents related work in Situation Awareness modeling. Section 4 describes the approach. In Section 5, the proposed methodology is explained. The paper concludes with Sections 6 and 7, which cover open issues and a final summary.

II. COGNITIVE MODELS OF HUMAN BEHAVIOR

To be able to profit from user data and at the same time avoid high costs and time consuming test procedures, cognitive architectures like ACT-R have shown to be an alternative way to the classic user testing methods. These models are able to simulate and predict human behavior, even in dynamic and complex environments. Initially such models replicated experiments conducted with humans in order to expand the knowledge about human cognition.

 ISBN: 978-1-61208-390-2

Today, cognitive models are able to produce valid predictions of human behavior even in complex and dynamic use cases like aviation and driving. For example, pilot [10] and driver models [11][12] gained a lot of attention in cognitive modeling. It was shown that the effects of devices like telephones and visual displays on the performance of the driving task can be simulated with a cognitive driver model and a traffic simulation [13][14].

The SA theory involves several cognitive processes. The theory builds upon these cognitive processes to describe how information directs operator performance. SA can help to explain when and why errors occur. However, the SA theory is not able to make predictions about operator performance. It is however possible to measure SA. State of the art methods include SAGAT (Situation Awareness Global Assessment Technique) [15], SPAM (Situation Present Assessment Method) [16] or SART (Situation Awareness Rating Technique) [17]. The SA measurement is often performed within a task simulation. These include driving simulations, flying simulations or air-traffic control simulations. SAGAT was introduced by Endsley to measure the SA of pilots. It is an offline measure where the simulation is stopped and questions about the situation are asked.

In contrast, the SPAM method is an online measure. The simulation does not have to be stopped to query operator SA. While the simulation is running, the operator is presented with a stimulus that indicates that they have to answer a situation related question. The operator decides when he will answer the question after the presentation of the stimulus. When the operator is able to listen and answer the query, he indicates that. Then, the question is asked while the simulation is running permanently. The time from the presentation of the question until the answer is here referred to as response time. Durso et al. state, that if the operator has "*in consciousness the information needed to answer a query*", response time is shorter [16]. Operators would still be able to answer correctly if they are able to search the display or environment for it. In that case, response time would be longer. It could be shown that the SPAM measures response time and accuracy have predictive power and are able to add to the incremental validity of a larger battery of cognitive tests [16].

For this work, the SPAM method itself, and results from existing studies where SPAM was applied, will contribute to the development of the model. Due to the following reasons this approach was chosen: First, SPAM offers performance measures for dynamic contexts. With latencies, such as response times, an important aspect of operator performance is evaluated because it guides how fast operators act. This is especially important for safety critical environments like driving. Second, SPAM is built to attribute to the dynamic characteristics of situations. SPAM does not interrupt the simulation, which underlines the dynamic aspect. Third, it can assess SA of the operators when "*it is successful, rather than only when SA fails*" [16]. These factors attribute to the applicability of SA measurements inside a cognitive architecture.

III. RELATED WORK

The SA theory gained a lot of attention and measurements of operator SA were developed and widely applied in various domains. SAGAT consists of a closed-loop simulation where (in that case) pilots fly a given scenario. At a random point in time, the simulation gets paused and the screen goes black. The pilot has to answer several questions (randomly selected from a larger set) concerning the situation to measure his knowledge. The answers of the pilot are compared to the aspects of the real situation to find out where the differences between the real and the perceived situation are. This method makes it possible to identify the SA elements pilots perceive and process depending on prior identified goals and tasks. Thus, it is possible to assess relevant knowledge of operators in a dynamic context. The development of SA evaluation methods also resulted in a use of this method in the industry where it is used to design new systems, train operators and measure the performance of operators to ensure optimal performance. SAGAT also got transferred to the driving domain where it was used in a variety of studies [18]. There has been work on driver reaction times to unexpected and expected road events, which revealed shorter brake response times for expected events [19].

Baumann and Krems propose that SA construction is comparable to language and text comprehension and state that "In both cases an integrated mental representation of the perceived and processed pieces of information is constructed." [20]. Their algorithmic approach to model SA aims to understand SA as a whole in order to extend the knowledge about the cognitive processes, which attribute to Endsley's initial theory [20]–[22]. Matthews [23] integrated driver's awareness of spatial, temporal, goal and system into a model of SA which is goal oriented and includes strategic, tactical and operational driving. The model aims to understand how modern intelligent transportation systems impact driver performance. Gugerty [24] used direct and indirect measures to assess driver's knowledge of the locations of other cars. This work provides implications about how people maintain SA in dynamic tasks like driving. In another work, Gugerty [25] reviews models and theories of attention, SA, comprehension and multitasking. Measures of SA are also presented. A collection of SA measurements is provided by Gawron where different techniques are presented [26].

IV. PROPOSED APPROACH

To be able to meet the objective of this work, it is planned to complete the following tasks:

1. Analysis of the fundamental cognitive processes which lead to variation in response and retrieval time of operators

 ISBN: 978-1-61208-390-2

2. Development of a theory of how these cognitive processes lead to variances in response times under consideration of continuous and event-based information techniques
3. Implementation of the theory from Step 2 in a cognitive architecture
4. Test of two design variants with the model and a driving simulation and predict response times to road and system events
5. Verification and model fit by an empirical evaluation of the model

It is planned to conduct the work in step 1-3 in two iterations. Starting from a first version, the model will be evaluated along the building process with human data. The question this work should answer is: How can the influence of continuous and event-based information on driver's reaction time to road / system events under partly automated driving be modeled with a cognitive architecture?

A. *Significance and Innovation*

Although SA helped to get insights about operator performance, to the knowledge of the author, no cognitive model incorporates SA as a foundation for the measurement of specific performance values like response times. Thus, this work extends the state of the art by proposing a method for the computational evaluation of assistance systems under the aspect of operator response time. To date, evaluations of SA are performed manually in complex settings with test personnel. Although the information gain with the existing methods is large, the applicability of these methods in the design process of assistant systems is limited. With cognitive models, design variants can be evaluated before actual user testing to identify presentation techniques, which are suitable for tests with users. Thus, the effort to evaluate such systems would decrease with the evaluation approach, this work offers.

To the knowledge of the author there are no models which allow an evaluation of driver reaction to external events under the consideration of information support from an assistance system. While the idea of system evaluation with cognitive models itself is not new, the approach of practically applying knowledge from existing SA rating techniques is novel and adds a valid contribution to the field. This is because models, which try to model SA in a cognitive architecture, assemble complex relationships between perception, memory and decision making in order to model a large amount of cognitive processes. The approach presented here is based on observations of existing test procedures and aims at the prediction of one specific performance measure (response time) as a resulting measure of SA. Thus, the complexity of the SA theory is limited to one factor, which makes the model building process less complex and manageable.

V. METHODOLOGY

The existing work introduced in the related work part of this proposal will be evaluated. The fundamental cognitive processes will be analyzed and based on this, a theory of how the response time under the influence of continuous and event-based information presentation are constructed, will be developed. This theory will be included in an existing driver model which will be the foundation for the development of the SA model. The driver model was built within the cognitive architecture CASCaS [10][27]. The CASCaS driver model consists of top-down visual attention mechanisms [27], bottom-up visual attention mechanisms are currently under development and will be integrated in the future. The symbolic representations of objects from the environment are transferred into the memory of the driver model. Concerning the Level 1 SA mechanisms, there is considered to be sufficient state of the art cognitive processes already implemented in the CASCaS architecture, so the perceptual part of the model will not be considered. The driver model will serve as a starting point for the exploration of how to include the cognitive processes, which will be developed in the model building process. Furthermore, a driving simulation is used where the cognitive driver model will be placed in. In such a closed-loop simulation, the model will be tailored to the use case in the platooning field. It will be supported by a symbolic representation of the two design variants which include continuous and event-based information presentation. In such a scenario, data will be generated from the model. In another step, the same scenario will be applied with human testers in a driving simulator. From the empirically gained data in this experiment, the model data will be compared to and the model fit will be calculated.

VI. QUESTIONS AND ISSUES

There are some issues which have to be considered. First, having response time as a dependent variable, it is important to examine and control the independent variables which lead to response time as a predictor of SA. Second, factors like experience, motivation and general cognitive capabilities attribute to driver performance. It is not clear at the moment how the interaction between these factors and the impact on performance measures like response time will add to the complexity of the model. Thus, for now it is assumed that these factors can be controlled by the study design.

VII. CONCLUSION

The presented dissertation project proposes a method to assess the impact of Ecological Interface Design variants on response times to road and system events of truck drivers. The foundation for this research includes current Situation Awareness measurements in dynamic contexts. The project extends the state of the art by using a specific performance measure (response time) as an indicator of SA inside a

 ISBN: 978-1-61208-390-2

cognitive architecture. Thus, the evaluation of driver assistance systems will be supported at design time.

ACKNOWLEDGMENTS

The research received funding from the European Commission Seventh Framework Program (FP7-ICT–2013-10) under Grant Agreement No. 610990, Project COMPANION. The authors thank Sebastian Feuerstack, Bertram Wortelen, Jan-Patrick Osterloh, Christian van Göns and Christian Denker for their helpful comments on earlier versions of this paper.

REFERENCES

[1] C. Bergenheim, S. Shladover, E. Coelingh, C. Englund, and S. Tsugawa, "Overview of platooning systems," Proceedings of the 19th ITS World Congress, 2012.

[2] M. Saffarian, J. C. F. de Winter, and R. Happee, "Automated Driving: Human-Factors Issues and Design Solutions," Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 56, no. 1, Oct. 2012, pp. 2296–2300.

[3] N. Sarter, D. Woods, and C. Billings, "Automation surprises," Handbook of human factors and ergonomics, vol. 2, 1997, pp. 1926–1943.

[4] M. Endsley, "Toward a Theory of Situation Awareness in Dynamic Systems," Human Factors: The Journal of the Human Factors and Ergonomics Society, vol. 37, no. 1, Mar. 1995, pp. 32–64.

[5] D. Kaber and M. Endsley, "Out-of-the-loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety," Process Safety Progress, vol. 16, no. 3, 1997, pp. 126–131.

[6] G. H. Walker, N. a. Stanton, T. a Kazi, P. M. Salmon, and D. P. Jenkins, "Does advanced driver training improve situational awareness?," Applied ergonomics, vol. 40, no. 4, Jul. 2009, pp. 678–87.

[7] K. J. Vicente and J. Rasmussen, "Ecological interface design: Theoretical foundations," IEEE Transactions on Systems, Man and Cybernetics, vol. 22, no. 4, 1992, pp. 589–606.

[8] C. M. Burns, G. Skraaning, G. a. Jamieson, N. Lau, J. Kwok, R. Welch, and G. Andresen, "Evaluation of Ecological Interface Design for Nuclear Process Control: Situation Awareness Effects," Human Factors: The Journal of the Human Factors and Ergonomics Society, vol. 50, no. 4, Aug. 2008, pp. 663–679.

[9] "COMPANION Project." [Online]. Available: http://www.companion-project.eu/.

[10] A. Lüdtke and J. Osterloh, "Simulating perceptive processes of pilots to support system design," Human-Computer Interaction–INTERACT, 2009, pp. 471–484.

[11] D. D. Salvucci, "Modeling Driver Behavior in a Cognitive Architecture," Human Factors: The Journal of the Human Factors and Ergonomics Society, vol. 48, no. 2, Jun. 2006, pp. 362–380.

[12] L. Weber, M. Baumann, A. Lüdtke, and R. Steenken, "Modellierung von Entscheidungen beim Einfädeln auf die Autobahn," 8. Berliner Werkstatt Mensch-Maschine-Systeme, 2009, pp. 86–91.

[13] J. Lee, J. D. Lee, and D. D. Salvucci, "Evaluating the distraction potential of connected vehicles," Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI, 2012, pp. 33–40.

[14] D. D. Salvucci, "Rapid prototyping and evaluation of in-vehicle interfaces," ACM Transactions on Computer-Human Interaction, vol. 16, no. 2, Jun. 2009, pp. 1–33.

[15] M. Endsley, "Situation awareness global assessment technique (SAGAT)," Proceedings of the IEEE 1988 National Aerospace and Electronics Conference, 1988, pp. 789–795.

[16] F. T. Durso, M. K. Bleckley, and A. R. Dattel, "Does Situation Awareness Add to the Validity of Cognitive Tests?," Human Factors: The Journal of the Human Factors and Ergonomics Society, vol. 48, no. 4, Dec. 2006, pp. 721–733.

[17] M. A. Vidulich and E. R. Hughes, "Testing a Subjective Metric of Situation Awareness," Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 35, 1991, pp. 1307–1311.

[18] N. Rauch, B. Gradenegger, and H. Krüger, "Die SAGAT-Methode zur Erfassung von Situationsbewusstsein im Fahrkontext," in Fortschritte der Verkehrspsychologie, vol. 1, J. Schade and A. Engeln, Eds. VS Verlag für Sozialwissenschaften, 2008, pp. 197–214.

[19] G. M. Fitch, M. Blanco, J. F. Morgan, and A. E. Wharton, "Driver Braking Performance to Surprise and Expected Events," in Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 2010, pp. 2076–2080.

[20] J. Krems and M. Baumann, "Driving and situation awareness: A cognitive model of memory-update processes," Human Centered Design, 2009, pp. 986–994.

[21] M. Baumann and J. Krems, "A comprehension based cognitive model of situation awareness," Digital Human Modeling, 2009, pp. 192–201.

[22] M. Baumann and J. F. Krems, "Situation awareness and driving: A cognitive model," in Modelling driver behaviour in automotive environments, Springer, 2007, pp. 253–265.

[23] M. Matthews and D. Bryant, "Model for situation awareness and driving: Application to analysis and research for intelligent transportation systems," Transportation Research Record: Journal of the Transportation Research Board, vol. 1779, 2001, pp. 26–32.

[24] L. Gugerty, "Situation awareness during driving: Explicit and implicit knowledge in dynamic spatial memory.," Journal of Experimental Psychology: Applied, vol. 3, no. 1, 1997, pp. 42–66.

[25] L. Gugerty, "Situation awareness in driving," in Handbook for driving simulation in engineering, medicine and psychology, 2011, pp. 19/1–19/8.

[26] V. J. Gawron, Human performance, workload, and situational awareness measures handbook. CRC Press, 2008.

[27] B. Wortelen, M. Baumann, and A. Lüdtke, "Dynamic simulation and prediction of drivers' attention distribution," Transportation Research Part F: Traffic Psychology and Behaviour, vol. 21, Nov. 2013, pp. 278–294.

 ISBN: 978-1-61208-390-2

Individual Differences in Deception and Deception Detection

Sarah Ita Levitan, Michelle Levine, Julia Hirschberg
Dept. of Computer Science
Columbia University
New York NY, USA
{sarahita,mlevine,julia}@cs.columbia.edu

Nishmar Cestero
Dept. of Psychology
Boston University
Boston MA, USA
nishmarc@bu.edu

Guozhen An, Andrew Rosenberg
Dept. of Computer Science
Queens College, CUNY
Queens NY, USA
{gan@gc,andrew@cs.qc}.cuny.edu

Abstract— **We are building a new corpus of deceptive and non-deceptive speech, using American English and Mandarin Chinese adult native speakers, to investigate individual and cultural differences in acoustic, prosodic, and lexical cues to deception. Here, we report on the role of personality factors using the NEO-FFI (Neuroticism-Extraversion-Openness Five Factor Inventory), gender, ethnicity and confidence ratings on subjects' ability to deceive and to detect deception in others. We report significant correlations for each factor with one or more aspects of deception. These are important for the study of trust, cognition, and multi-modal information processing.**

Keywords-Deception; cross-linguistic; personality.

I. INTRODUCTION

Finding new methods for detecting deception is a major goal of researchers in psychology and computational linguistics as well as commercial, law enforcement, military, and intelligence agencies. While many new techniques and technologies have been proposed and some have even been fielded, there have been few significant successes. The goal of our research is to develop techniques to identify deceptive communication in spoken dialogue. As part of this investigation, we are focusing on how within-culture and cross-cultural differences between deceivers as well as their common characteristics impact deceptive speech behavior. Our research focuses solely on cues drawn from the speech signal, which have been little studied.

In this paper, we describe results of experiments correlating gender, ethnicity, and personality characteristics from the NEO-FFI Five Factor Analysis [1] with subjects' ability to deceive and to judge deception in others' speech. We also examine the importance of subjects' reported confidence in their judgments in deception production and detection. In Section 2, we describe previous work on cues to deception and deception detection. In Section 3 we discuss our experimental design, data collection and annotation. In Section 4 we describe results of our correlations of personality, gender, ethnicity, and confidence on deception production and detection. We conclude in Section 5 with a discussion of our results.

II. DECEPTION DECTECTION

Previous research on deceptive behavior has studied standard biometric indicators commonly measured in polygraphy (cardiovascular, electrodermal, and respiratory), facial expression, body gestures, brain imaging, body odor, and lexical and acoustic-prosodic information.

Biometric measures are widely acknowledged, even by polygraphers, to be inadequate for deception detection, performing at no better than chance. Useful groundwork has been laid in identifying potential facial expression cues of deception by Ekman et al. [2][3]. However, attempts to identify deception from facial expressions are questioned by some researchers [4]. Moreover these approaches are difficult to automate, requiring delicate image capture technology and laborious human annotation. There have been promising results using automatic capture of body gestures as cues to deception [5][6], but this method requires multiple, high-caliber cameras to capture movements reliably. Similarly, the use of brain imaging technologies for deception detection is still in its infancy [7] and these require the use of MRI techniques, which are not practical for general use. Additional biometric indicators of deception such as body odor are beginning to be investigated [8] but these studies, like brain imaging, are in very early stages.

Some researchers and practitioners have examined language-based cues to deception. These include Statement Analysis [8], SCAN [10][11], and some of the text-based signals identified by John Reid and Associates [12]. These efforts have been popular among law enforcement and military personnel, though little tested scientifically (although Bachenko et al. [13] have partially automated and validated some features used in Statement Analysis). Other lexical cues to deception have been developed and tested empirically by Pennebaker and colleagues [14][15] and by Hancock et al. [16]. There has also been research focused on lexical cues to deception in written online communication [17][18].

Little work has been done on cues to deception drawn from the speech signal. Simple features such as intensity and hypothesized vocal tremors have performed poorly in objective tests [19][10][21][22], although other features examined by Harnsberger et al. [23] and Torres et al. [24] have had more success. In previous work on deception in American speech, Hirschberg et al. [25] developed automatic deception detection procedures trained on spoken cues and tested on unseen data. These procedures have achieved accuracies 20% better than human judges. In the process of identifying common characteristics of deceivers, they also

 ISBN: 978-1-61208-390-2

noticed a range of individual differences in deceptive behavior, e.g., some subjects raised their pitch when lying, while some lowered it significantly; some tended to laugh when deceiving, while others laughed more while telling the truth. They also discovered that human judges' accuracy in judging deception could be predicted from their scores on the NEO-FFI, suggesting that such simple personality tests might also provide useful information in predicting individual differences in deceptive behavior itself [26].

Differences in verbal deceptive behavior in different cultures have been identified by several researchers [27][28]. Studies of deceptive behavior in non-Western cultures have primarily focused on understanding how culture affects *when* people deceive and *what* they consider deception [29][30]. Studies investigating the universality of deceptive behavior have found that, while stereotypes may exist [31] these may not correlate with actual deceptive behavior [32][33] and that culture-specific deception cues do exist [27][28][34].

In the work presented here, we investigate both the ability to deceive and to detect deception considering gender and ethnicity and examining new cues to deception: features extracted from the NEO-FFI personality inventory [1] and subjects' reported confidence in their abilities.

A. Experimental Design

To investigate questions of individual and cross-cultural differences in deception perception and production, we are collecting a large corpus of cross-cultural deceptive and non-deceptive speech. We employ a variant of the 'fake resume' paradigm to elicit both deceptive and non-deceptive speech from native speakers of Standard American English (SAE) and Mandarin Chinese (MC), both speaking in English. Each conversation in the corpus is between a pair of subjects who are not previously acquainted with one another. To date, the corpus includes 134 conversations between 268 subjects.

For the first phase of each session, subjects are separated from one another. Each is told that they will play a lying game with another subject, in which they will alternate between interviewing their partner and being interviewed themselves. As interviewees, they should attempt to successfully deceive the interviewer. As interviewers, they should attempt to determine whether the interviewee is lying or telling the truth. For motivation, they are told that their compensation depends on their ability to deceive while being interviewed, and to judge correctly while interviewing. As interviewer, they receive $1 each time they correctly identify an interviewee's answer as either lie or truth and lose $1 for each incorrect judgment. As interviewee, they earn $1 each time their lie is judged to be true, and lose $1 each time their lie is correctly judged to be a lie by the interviewer.

Subjects are then asked to truthfully complete a 24-item biographical questionnaire. In addition to their true answers, they are told to create a false answer for a random half of the questions. They are given guidelines to ensure that their false answer differed significantly from the truth, to ensure that lying will not be too easy. For example, for the question "Where were you born," the false answer must be a place that the subject has never visited, a false answer to "What is your father's occupation" must be different from their mother's true occupation, and so on. Before the interviews begin, false answers are checked by an experimenter to make sure subjects follow these guidelines. In addition to the biographical questionnaire, each subject completes the NEO-FFI personality inventory [1], which is described below.

While one subject is completing the NEO-FFI inventory, we collect a 3-4 minute baseline sample of speech from the other participant for use in speaker normalization. The experimenter elicits natural speech by asking the subject open-ended questions (e.g., "What do you like best/worst about living in NYC?"). Subjects are instructed to be truthful during this part of the experiment. Once both subjects have completed all the questionnaires and we have collected baseline samples of speech, the lying game begins.

The lying game takes place in a sound booth where the subjects are seated across from each other, separated by a curtain so that there is no visual contact; this is necessary since our focus is on spoken and not visual cues. There are two parts to each session. During the first half, one subject acts as the interviewer while the other answers the biographical questions, lying for half and telling the truth for the other half, based on the modified questionnaire. In the second part of the session, the subjects switch roles. All speech data is collected in a double-walled sound booth in the Columbia Speech Lab and recorded to digital audio tape on two channels using Crown CM311A Differoid head-worn close-talking microphones.

The interviewer is able to ask the questions in any order s/he chooses, and is encouraged to ask follow-up questions to help determine the truth of the interviewee's answers. For each question, the interviewer records his/her judgment, along with a confidence score from 1-5. As the interviewee answers the questions, s/he presses a T or F key on a keyboard (which the interviewer cannot see) for each phrase, logging each segment of speech as true or false. Thus, while the biographical questionnaire provides the 'global' truth value for the answer to the question asked, the key log provides the 'local truth' value for each phrase, which is automatically aligned with each speech segment. At the end of the experiment, subjects complete a brief questionnaire, which includes additional confidence questions.

B. Personality Assessment

The NEO-FFI personality assessment [1] is based on the five-factor model of personality, an empirically-derived and comprehensive taxonomy of personality traits. It was developed by applying factor analysis to thousands of descriptive terms found in a standard English dictionary. It is used to assess the five personality dimensions of:

Openness to Experience. Designed to capture imagination, aesthetic sensitivity, and intellectual curiosity. It is "related to aspects of intelligence, such as divergent thinking, that contribute to creativity" [1]. Those who score low on this dimension prefer the familiar and tend to behave more conventionally. People high in Openness are "willing to entertain novel ideas and unconventional values" [1].

 ISBN: 978-1-61208-390-2

Conscientiousness. Addresses individual differences in self-control, such as the ability to control impulses, but also to plan and carry out tasks. It measures contrasts between determination, organization, and self-discipline and laxness, disorganization, and carelessness.

Extraversion. Meant to capture proclivity for interpersonal interactions, and variation in sociability. It reflects contrasts between those who are reserved vs. outgoing, quiet vs. talkative, and active vs. retiring.

Agreeableness. Measures interpersonal tendencies and is intended to assess an individual's fundamental altruism. Individuals high in Agreeableness are sympathetic to others and expect that others feel similarly.

Neuroticism. Contrasts emotional stability with maladjustment. It is intended to capture differences between those prone to worry vs. calm, emotional vs. unemotional behavior, and vulnerable vs. hardy.

III. ANALYSES AND RESULTS

Although subjects were instructed to lie in response to 12 of the questions, 55 out of 268 subjects did not follow these instructions, and lied in response to more or fewer than 12 questions. The following analyses include 126 pairs, those in which both subjects lied in response to 10-14 of the questions; this restriction ensures that roughly equal amounts of truthful and deceptive speech are available for each subject. This subsample consists of 142 native SAE participants (88 females, 54 males) and 110 native MC participants (69 females, 41 males). Our eventual goal is a corpus balanced for gender and ethnicity, but in this paper we present results only on this sample.

First, we examined how accurately subjects could identify deception in their partners during the lying game. Prior research indicates that human judges perform worse than chance at detecting deception [26][35]. However, in our study subjects correctly identified question responses as truthful or deceptive at a greater than chance level. They were accurate 56.75% of the time (compared to the chance baseline of 49.55%).

To further assess subjects' accuracy, we explored how well subjects detected lies as opposed to truths. To account for the different number of lies across subjects, for each subject we calculated ratio scores for: number of successful global lies to the number of global lies told (*successful lies*); the number of successful lie detections to the number of global lies told (*successful lie detections*); the number of successful truth detections to the number of truths told (*successful truth detections*). Results indicate that people successfully deceived their partner 51.83% of the time. Deceptive answers were correctly identified 48.16% of the time and truthful answers were correctly identified 65.20% of the time.

We investigated whether subjects' ability to detect deception was correlated with their ability to deceive by comparing *successful lies* to *successful lie detections*. Our data indicate that subjects who were better at detecting deceptive answers were also better at deceiving, $r(252) = 0.13$, $p = 0.04$. When separated by gender and native language, it becomes apparent that this correlation is strongest for females, and specifically for SAE females ($r(157) = 0.24$, $p = 0.003$ and $r(88) = 0.29$, $p = 0.005$). We note that, for all subjects, those who were better at detecting deception were also more likely to label their partners' answers as untrue --- whether or not their partner did indeed lie, $r(252) = 0.69$, $p < 0.001$. However, female subjects who were more likely to label their partners' answers lies *were* also better at deceiving, $r(157) = 0.18$, $p = 0.02$.

Next, we examined how individual differences in gender, culture, personality, and confidence ratings interacted with successful deception and deception detection. Independent sample t-tests indicated no effect of subjects' gender or native language on their ability to deceive. In addition, correlational analyses showed no effect of personality factors on subjects' ability to detect deception. This latter finding is in sharp contrast with Enos et al.'s findings for personality differences in success rates of post hoc judges of deception [24] and suggests that personality factors may play a more important role when non-conversational participants rather than those engaged in the conversation are judging deception. In contrast, the personality factor of Extraversion does correlate with subjects' ability to deceive and here we do find cultural and gender differences: MC females' success positively correlates with Extraversion scores ($r(69) = 0.26$, $p = .03$) while SAE males' success *negatively* correlates with their Extraversion scores ($r(54) = -0.36$, $p = .01$). Furthermore, SAE females' deception ability *negatively* correlates with their Conscientiousness scores ($r(86) = -0.22$, $p = .04$).

For confidence ratings, we also find a gender difference: overall, female subjects' ability to detect deception *negatively* correlates with their average confidence in their judgments, $r(157) = -0.20$, $p = 0.01$. This did not hold true for SAE females examined separately although it did for MC females, $r(69) = -0.26$, $p = 0.03$. We hypothesize that interviewers who are less confident in their judgments may ask more follow-up questions and thus obtain more evidence to determine deception. It will be important to look at answer length and number of follow-up questions to test these possibilities. We also found that, for females, average confidence in detecting deception *negatively* correlated with Neuroticism, $r(155) = -0.16$, $p = 0.05$. Not surprisingly, women who are less "neurotic" are more confident in their deception judgments. We will need to check for similar findings for male subjects once we have collected more data.

Finally, we looked at whether the gender and culture of subjects' partners played a role in deception and deception detection. Independent t-tests show no effects so far.

IV. CONCLUSIONS AND FUTURE WORK

Preliminary analysis of a sample of our deceptive speech corpus shows some promising results: We found that subjects who are better at detecting lies are also better at deceiving others, and that this correlation is stronger for females and stronger still for SAE females. While we have not found effects of personality characteristics on our subjects' ability to detect deception, in contrast to Enos et al.

 ISBN: 978-1-61208-390-2

[26], we have found that Extraversion and Conscientiousness scores correlate with ability to deceive, although the direction of this effect differs depending upon gender and ethnicity. We note the difference between the judgment tasks in our experiment vs. [26]'s. Finally, we found that MC women showed a negative correlation between confidence scores and ability to detect deception while SAE women and men in general did not. In addition, for all females, ability to detect deception was negatively correlated with Neuroticism.

We anticipate that the completion of a balanced corpus will clarify and expand some of these findings. We will also finish transcribing our corpus and aligning the transcription with the speech recordings so that we can add acoustic, prosodic, and lexical cues to gender, ethnicity, and personality information for the purpose of building automatic classifiers for deceptive vs. non-deceptive speech.

V. ACKNOWLEDGMENTS

We thank the following students for their contributions to this study: Zoe Baker-Peng, Ling Huang, Melissa Kaufman-Gomez, Yvonne Missry, Elizabeth Petitti, Sarah Roth, Molly Scott, Jennifer Senior, Grace Ulinski, and Christine Wang. This work was partially funded by AFOSR FA9550-11-1-0120.

REFERENCES

[1] P. T. Costa and R. R. McCrae, "Revised NEO personality inventory (NEO PI-R) and NEO five-factor inventory (NEO-FFI)." Odessa, FL: Psychological Assessment Resources, 1992.

[2] P. Ekman, M. Sullivan, W. Friesen, and K. Scherer, "Face, voice, and body in detecting deception," Journal of Nonverbal Behaviour, vol. 15(2), 1991, pp. 125-135..

[3] M. Frank, M. O'Sullivan, and M. Menasco, "Human behavior and deception detection," in J. G. Voeller (Ed.), Wiley Handbook of Science and Technology for Homeland Security, New York: John Wiley & Sons, 2008.

[4] J. Cohn quoted in http://abcnews.go.com/GMA/HealthyLiving/autism-research-benefit-studying-babies-facial-recognition-experts/story?id=9244817&page=3, 12/04/2009.

[5] T. Qin, J. Burgoon, and J. Nunamaker, "An exploratory study on promising cues in deception detection and application of decision tree." Hawaii International Conference on System Sciences, 2004, pp. 23-32.

[6] T. Meservy et al., "Deception Detection through Automatic, Unobtrusive Analysis of Nonverbal Behavior." IEEE Intelligent Systems, 20:5 (September 2005), pp. 36-43.

[7] D. Langleben et al., "Telling truth from lie in individual subjects with fast event-related fMRI." Human Brain Mapping, 2005, 26(4): 262-72.

[8] S. Waterman, "DHS wants to use human body odor as biometric identifier, clue to deception," United Press International (UPI). http://www.upi.com/Top_News/Special/2009/03/09/DHS-wants-to-use-human-body-odor-as-biometric-identifier-clue-to-deception/UPI-20121236627329/, 3/9/2009.

[9] S. Adams, "Statement analysis: What do suspects' words really reveal?" FBI Law Enforcement Bulletin, October 1996.

[10] A. Sapir, "Scientific Content Analysis (SCAN)." Laboratory of Scientific Interrogation. Phoenix, AZ, 1987.

[11] N. Smith, "Reading between the lines: An evaluation of the scientific content analysis technique (SCAN)," Police Research Series. London, UK, 2001.

[12] J. E. Reid and Associates, "The Reid Technique of Interviewing and Interrogation," Reid, John E. and Associates, Inc., 2000.

[13] J. Bachenko, E. Fitzpatrick, and M. Schonwetter, "Verification and implementation of language-based deception indicators in civil and criminal," International Conference on Computational Linguistics, vol. 1, Manchester, 2008, pp. 41-48.

[14] Pennebaker, M. Francis, and R. Booth, "Linguistic Inquiry and Word Count." Erlbaum Publishers, Mahwah, NJ, 2001.

[15] M. Newman, J. Pennebaker, D. Berry, and J. Richards, "Lying words: Predicting deception from linguistic style." Personality and Social Psychology Bulletin, 2003, 29:665-675.

[16] J. Hancock, L. Curry, S. Goorha, and M. Woodworth, "On lying and being lied to: A linguistic analysis of deception." Discourse Processes, vol. 45, pp. 1-23, 2008.

[17] L. Zhou and D. Zhang, "Following Linguistic Footprints: Automatic Deception Detection in Online Communication." Communications of the ACM, 2008, 51(9): 119-122.

[18] R. Mihalcea and C. Strapparava, "The lie detector: explorations in the automatic recognition of deceptive language." In Proceedings of the ACL-IJCNLP 2009. Stroudsburg, PA.

[19] D. Haddad and R. Ratley, "Investigation and evaluation of voice stress analysis technology." http://www.ncjrs.org/pdffiles1/nij/193832.pdf, March 2002.

[20] H. Hollien and J. Harnsberger, "Voice stress analyzer instrumentation evaluation." Technical report, Counterintelligence Field Activity. 2006.

[21] H. Hollien, J. Harnsberger, C. Martin, and K Hollien, "Evaluation of the NITV CVSA." Journal of Forensic Sciences, vol. 53, pp. 183 – 193, 2006.

[22] A. Eriksson and F. Lacerda, "Charlatanry in forensic speech science: A problem to be taken seriously." The International Journal of Speech, Language and the Law, vol. 14:2, pp. 169-193, 2007, http://www.scribd.com/doc/9673590/Eriksson-Lacerda-2007.

[23] J. Harnsberger, H. Hollien, C. Martin, and K. Hollien, "Stress and deception in speech: evaluating layered voice analysis." Journal of Forensic Sciences, vol. 54(3), pp. 642-50, 2009.

[24] J. Torres, E. Moore, and E. Bryant, "A Study of Glottal Waveform Features for Deceptive Speech Classification." ICASSP 2008, Las Vegas.

[25] J. Hirschberg et al., "Distinguishing deceptive from non-deceptive speech." Interspeech 2005. Lisbon.

[26] F. Enos, S. Benus, R. Cautin, M. Graciarena, J. Hirschberg, and E. Shriberg, "Personality factors in human deception detection: Comparing human to machine performance." Interspeech 2006. Pittsburgh.

[27] F. Feldman, "Nonverbal disclosure of deception in urban Koreans." Journal of Cross-Cultural Psychology, vol. 10, 1979, pp. 215-221.

[28] M. Cody, W. Lee, and E. Y. Chao, "Telling lies: Corellates of deception among Chinese." In J.P. Forgas and J. M. Innes, eds. Recent Advances in Social Psychology: An International Perscpective, 1989, pp.359-568.

[29] M. Lapinski and T. Levine, "Culture and information manipulation theory: The effects of self-construal and locus of benefit on information manipulation." Communication studies, vol. 51(1), 2000, pp. 55-73.

[30] J. Seiter, J. Bruschke, and C. Bai, "The acceptability of deception as a function of perceivers' culture, deceiver's

 ISBN: 978-1-61208-390-2

intention, and deceiver-deceived relationship." Western Journal of Communication 66(2), 2002, pp.158-180.

[31] C. Bond and The Global Deception Research Team, "A world of lies." Journal of Cross-Cultural Psychology, vol. 37(1), 2006, pp. 60-74.

[32] A. Vrij and G. Semin, "Lie experts' beliefs about nonverbal indicators of deception." Journal of Nonverbal Behavior, vol. 20(1), 1996, pp. 65-80.

[33] M. Zuckerman, B. DePaulo, and R. Rosenthal, "Verbal and non-verbal communication of deception." In L. Berkowitz (ed.), Advances in Experimental Social Psychology, Academic Press, New York, pp.1-59, 1981.

[34] R. Feldman, L. Jenkins, and O. Popoola, "Detection of deception in adults and children via facial expressions." Child development, 350-355,1979).

[35] M. Aamondt and H. Custer, "Who can best catch a liar?" Forensic Examiner 15(1):6-11. 2006.

 ISBN: 978-1-61208-390-2

Automatic face recognition using SIFT and networks of tagged neural cliques

Ehsan Sedgh Gooya, Dominique Pastor and Vincent Gripon

Institut Mines Telecom; Telecom Bretagne; UMR CNRS 6285 Lab-STICC
Email: name.surname@telecom-bretagne.eu

***Abstract*—Bearing information by a fully interconnected subgraphs is recently improved in the neural network of cliques. In this paper, a face recognition system is presented using such networks where local descriptors are used to perform feature extraction. In the wide range of possible image descriptors for face recognition, we focus specifically on the Scale Invariant Feature Transform (SIFT). In contrast to standard methods, our proposed method requires no empirically chosen threshold. Moreover, it performs matching between sets of features, in addition to individual feature matching. Thus, we favor joint occurrences of descriptors during the recognition process. We compare our approach to state of the art face recognition systems based on SIFT descriptors. The evaluation is carried out on the Olivetti and Oracle Research Laboratory (ORL) face database, whose diversity is significant for assessing face recognition methods.**

Keywords–Face recognition, neural networks, associative memories, neural cliques, SIFT descriptors.

I. INTRODUCTION

Instantly recognizing a familiar face is easy task for humans. However, as many processes related to vision, automatic pattern recognition is generally difficult. Face recognition is among the most visible and challenging research topics in computer vision and automatic pattern recognition [1], and many methods, such as Eigenfaces [2], Fisherfaces [3] and SVM [4], have been proposed in the past two decades. Recently the sparse representation (or coding) based classication (SRC) has been successfully used in face recognition [5], [6]. In SRC, the testing image is represented as a sparse linear combination of the training samples, and the representation delity is measured by the $l_1 - norm$ of coding residual.

However, the last word in pattern matching is the human brain in the sense that it seeks to identify links between what it currently observes and what it has experienced in the past.

Over the last ten years, much attention has been given to feature-based methods such as SIFT [7]. This is due to the fact these descriptors remain invariant under rotation, scaling and variation in lightning condition. In a conventional method, SIFT features are extracted from all the faces in the database. Then, given a query face image, each feature extracted from that face is compared to those of each face in the database. A query feature is considered to match one of the database according to a certain threshold-based criterion. The face in the database with the largest number of matched descriptors is considered as the nearest face.

Although the nearest face criterion may give very good results, it suffers from the following limitations. To begin with, only the first nearest neighbors are used to characterize the contents of the database. Also, the threshold set by the user is obtained a posteriori and, as such, varies from one experiment to another.

To overcome these drawbacks, we propose a novel approach based on matching sets of descriptors. This approach relies on a new extension of the neural network introduced in [8] and [9] that embeds messages to learn into cliques. Basically, this neural network is an associative memory (denoiser). However, to the best of our knowledge, it is the first time that it is used for pattern recognition.

The reason why we investigate using clique networks with SIFT descriptors is because of their error correcting capability. Intuitively, the mismatches that may occur when pairing SIFT descriptors may be corrected by the redundancy of clique patterns in the neural network.

The face recognition method proposed in this paper combines SIFT features and networks of neural cliques. In the course of the paper, the fundamental concepts of the SIFT algorithm are presented in Section II. The third section (Section III) reviews different sift matching methods for face recognition. The neural network of neural cliques is described in Section IV and the whole face recognition system based on neural network of neural cliques is presented in Section V. Finally, we present and discuss the results of the proposed face recognition system in Section VI.

II. SCALE INVARIANT FEATURE TRANSFORM (SIFT)

The Scale Invariant Feature Transform algorithm was proposed by David G. Lowe in [7] and extracts distinctive features. These features are invariant to rotation, scaling and partly invariant to changes in illumination and affine transformation of images. Therefore, these features are good candidates for face recognition. The main steps to calculate the SIFT features of an image are the following ones.

A. Keypoint localization

To efficiently detect stable keypoint locations, scale-space extrema in the difference-of-Gaussian (DoG) are used during the computation of the SIFT descriptors. The scale-space is defined as a function $L(x, y, \sigma)$ obtained by Gaussian kernel convolution with the input image so that:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \tag{1}$$

where $I(x, y)$ is the input image and $G(x, y, \sigma)$ is the Gaussian function:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2}\, e^{-\frac{x^2+y^2}{2\sigma^2}} \ . \tag{2}$$

 ISBN: 978-1-61208-390-2

$DoG\,(x, y, \sigma)$ can be computed from the difference of two nearby scales separated by an empirically chosen constant multiplicative factor k:

$$DoG(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ = L(x, y, k\sigma) - L(x, y, \sigma)\,. \quad (3)$$

The efficient approach to construction of $DoG(x, y, \sigma)$ is shown in Figure 1.

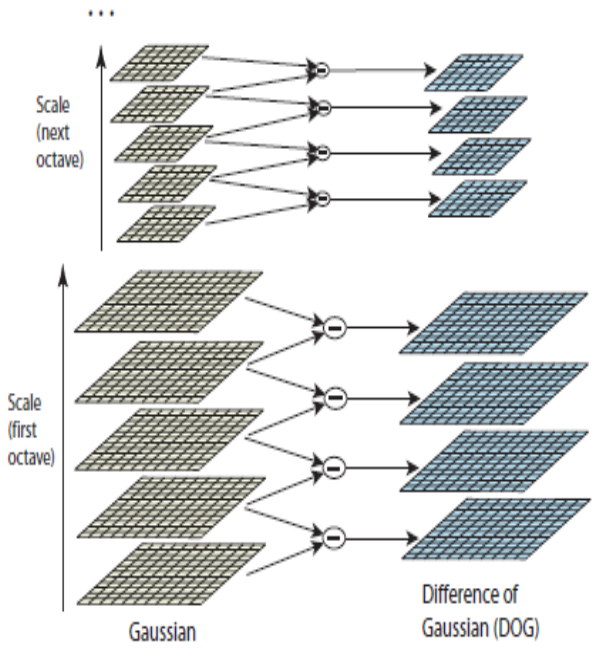


Figure 1. Excerpt from [7]). The convolved images are grouped by octave (an octave corresponds to doubling the value of σ), and the value of k_i is selected so that we obtain a fixed number of convolved images per octave. Then the Difference-of-Gaussian (DoG) images are taken from adjacent Gaussian-blurred images per octave.

In order to detect the local maxima and minima of $DoG(x, y, \sigma)$, each sample point is compared to its eight neighbors in the current image and to its nine neighbors in the scales above and below, as shown in Figure 2. After comparison, the sample is selected only if it is larger than all of these neighbors or smaller than all of them. Moreover, the algorithm eliminates candidates that are located on an edge or have poor contrast.

B. Assigning Rotation to Keypoint

Given a keypoint at position (x_0, y_0) for a given scale σ_0, the gradient principal direction must be computed. To do so, for each pixel (x, y) directly connected to x_0, y_0 we compute the magnitude $m(x, y)$ and orientation $\theta(x, y)$ as follows:

$$m(x, y) = \\ \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \\ \theta(x, y) = \tan^{-1} \frac{L(x, y-1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)}\,. \quad (4)$$

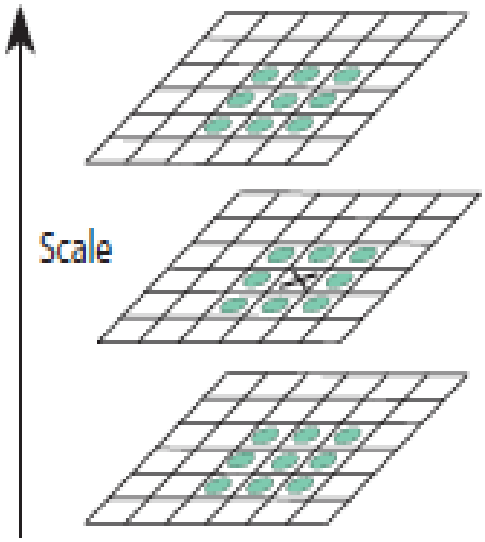


Figure 2. (Excerpt from [7]). Maxima and minima of the DoG images are detected by comparing a pixel (marked with X) to its 26 neighbors in 3×3 regions at the current and adjacent scales (marked with circles).

An orientation histogram with 36 bins covering 360 degrees is formed from the gradient orientations of the sample points around the keypoint. Each sample added to the histogram is weighted by its gradient magnitude. Then, the maximum orientation is assigned to this keypoint. For any other orientation within 80% of the maximum orientation, a new keypoint is created with this orientation. Each keypoint is rotated in direction of its orientation and then normalized. The maximum orientation, θ_0, is assigned to the keypoint. A keypoint is then entirely determined given the four parameters $(x_0, y_0, \sigma_0, \theta_0)$.

C. Construction of the feature descriptors

The 4×4 subregions located around a given keypoint are delimited, each containing 4×4 pixels. In each subregion, the orientations and magnitudes at each pixel are calculated. An orientation histogram of 8 bins is computed for each subregion. The corresponding gradient values are weighted by a Gaussian circular window. The 16 resulting histograms are then normalized and form a vector with 128 dimensions (16×8).

Figure 3 shows an example of a keypoint with its descriptor and orientation.

III. Review of SIFT-based matching methods for face recognition

A. Aly's matching

In [10], each SIFT descriptor in the test image is compared with every descriptor of each training image. The comparison is performed using cosine similarity of two feature $\mathbf{f_1}$ and $\mathbf{f_2}$ computed as follows:

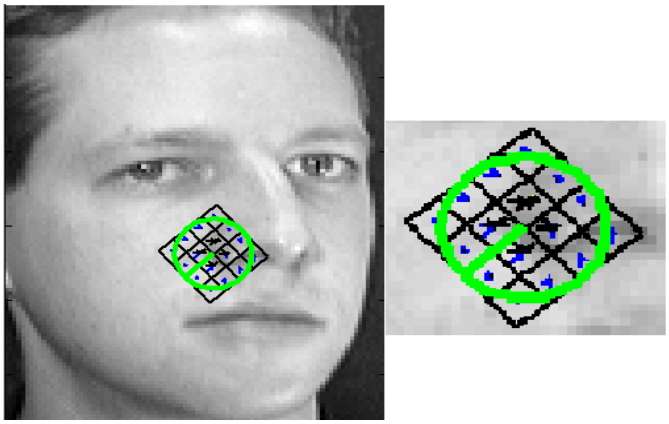


Figure 3. Example of a SIFT descriptor and orientation.

$$\Omega(\mathbf{f_1}, \mathbf{f_2}) = \frac{\mathbf{f_1} \times \mathbf{f_2}}{\|\mathbf{f_1}\| \times \|\mathbf{f_2}\|} \ . \tag{5}$$

A feature $\mathbf{f}$ from the query face image is considered to match a feature $\mathbf{f_1}$ from the gallery images if a) $\mathbf{f_1}$ is the most similar feature to $\mathbf{f}$ in the gallery images and b) the second closest feature $\mathbf{f_2}$ to $\mathbf{f}$ in the gallery images is such that $\Omega(\mathbf{f_2}, \mathbf{f}) - \Omega(\mathbf{f_1}, \mathbf{f}) \geq \Omega_{\min}$, where $\Omega_{\min}$ is a fixed threshold.

B. Lenc-Kral and Kepenekci's matching

In [11], for each feature of the query face image, the most similar feature of the gallery face is identified. The sum of the highest similarities is computed and is used as a measure of similarity between two faces. Kepenekci's SIFT matching combines two methods of matching and uses a weighted sum of the two values as a result. The cosine similarity is employed for feature comparison.

C. Support Vector Machine classifier

The face recognition method presented in [12] employs SIFT features to extract discriminative local features and Support Vector Machine (SVM) as a classifier. Basically, SVM is able to separate positive and negative examples using decision surfaces constructed by optimal separating hyperplanes.

IV. NETWORKS OF TAGGED NEURAL CLIQUES

An associative memory is a device capable of storing vectors, then retrieving them when some coordinates are missing or altered. Recently, an implementation of an associative memory based on neural networks was proposed in [8]. This network can store a large number of binary vector patterns and retrieve them with low error probability and high memory efficiency, even in case of erasures. The principle of this model is to embody vectors into fully interconnected subgraphs called cliques. Contrary to the celebrated Hopfield model [13], connections are binary in [8].

However, the vectors that can be handled by such a network are too constrained for our application in face recognition. For this reason, we follow the extension proposed in [9], in which any binary vector can be handled by the network. We extend the functionalities of this model so as to perform classification of vector patterns for face recognition.

As any associative memory, two operations are performed by the network: storing and retrieving. In the following subsections, we describe these two operations.

A. Storage (learning) process

In this paper, the binary neural network contains n neurons. This network stores c gallery patterns. Each gallery pattern $\mathbf{g}_k$ is defined as the concatenation between pattern vector $\mathbf{x}_k$ with dimension d and its associated class represented by vector $\mathbf{e}_k$ (k-th element of the canonical base in $\mathbb{R}^{\mathrm{c}}$):

$$\mathbf{g}_k = \begin{pmatrix} \mathbf{x}_k \\ \mathbf{e}_k \end{pmatrix} \in \{0;1\}^{d+c} \ . \tag{6}$$

Let us index these neurons from 1 to $n = d + \mathrm{c}$. The underlying graph is fully represented by its adjacency binary matrix $\boldsymbol{W}$ of size $n \times n$, in which the gallery patterns are stored.

In the storage procedure, the first d neurons are employed to embody the vectors $\mathbf{x}_k$ in the form of cliques and the c remaining neurons are used to tag the cliques (see figure 4). Accordingly, $\mathbf{e}_k$ is the binary unit vector whose unique coordinate equal to 1 is the tag index associated with $\mathbf{x}_k$. Denote by $(\mathbf{g}_k)_{1 \leq k \leq \mathrm{c}}$ the sequence of gallery vectors to be stored. Then $\boldsymbol{W}$ is defined as:

$$\boldsymbol{W} = \max_k \left(\mathbf{g}_k \cdot \mathbf{g}_k^{\mathrm{T}} \right) \tag{7}$$

where $(\bullet)^{\mathrm{T}}$ is the transpose operator. Using this process, the connection between neurons i and j is set to 1 if there exists k such that $\mathbf{g}_k(i) = \mathbf{g}_k(j) = 1$.

According to the foregoing, the computation of the adjacency matrix $\boldsymbol{W}$ is independent of the order in which gallery vectors are presented. Moreover, adding a new gallery pattern can be done online, independently of previously stored patterns.

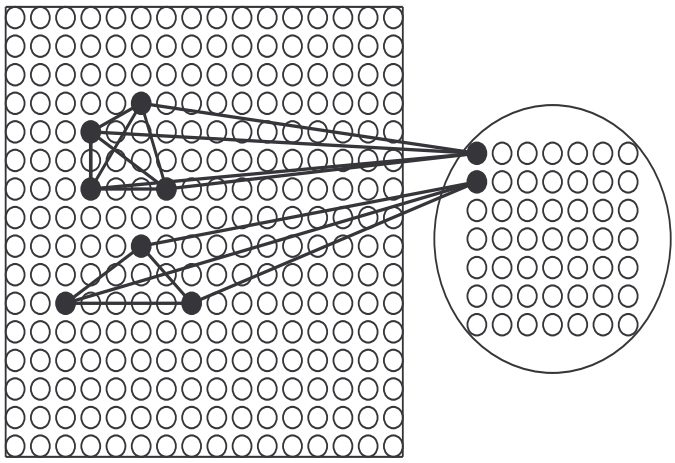


Figure 4. Example of two tagged cliques: the first d neurons on the left are used to embody the vectors $\mathbf{x}_k$ in the form of cliques and the c neurons on the right are employed to tag the cliques.

 ISBN: 978-1-61208-390-2

B. Retrieving process

The retrieving algorithm is a two-step and possibly iterative procedure. The first step aims at matching the input with a similar clique in the neural network. The purpose of the second step is to retrieve the associated label. To perform retrieval, we use a nonlinear filter f that operates over a vector $\mathbf{v}$. It consists in putting to zero all the coordinates that are not maximum and to one those that reach the maximum:

$$f(\mathbf{v})_i = \begin{cases} 1 & \text{if } \mathbf{v}_i = \max_j \mathbf{v}_j \\ 0 & \text{otherwise} \end{cases} . \quad (8)$$

To ease the reading, we also introduce the operator $\pi_{\mathbf{x}}(\mathbf{v})$ (resp. $\pi_{\mathbf{e}}(\mathbf{v})$) that extracts the vector made of the first (resp. last) d (resp. c) coordinates of $\mathbf{v}$. Conversely, we denote by $(\mathbf{v};\mathbf{v}')$ the concatenation of vectors $\mathbf{v}$ and $\mathbf{v}'$. Finally, we denote by $\mathbf{0}^{\mathrm{c}}$ the zero vector with dimension c. Algorithm 1 is used to classify $\mathbf{x}$.

Algorithm 1: Classification algorithm with neural network of tagged cliques.

Input: Input pattern $\mathbf{x}$ and adjacency matrix $\boldsymbol{W}$
Output: $\widehat{\mathbf{e}}_k$, the class indicator vector estimated for $\mathbf{x}$

1 $\widehat{\mathbf{x}} = \pi_{\mathbf{x}}\left(f(\boldsymbol{W}\begin{pmatrix}\mathbf{x}\\ \mathbf{0}^{\mathrm{c}}\end{pmatrix})\right)$;

2 $\widehat{\mathbf{e}}_k = \pi_{\mathbf{e}}\left(f(\boldsymbol{W}\begin{pmatrix}\widehat{\mathbf{x}}\\ \mathbf{0}^{\mathrm{c}}\end{pmatrix})\right)$;

If the output is not a unit vector then we consider that the classification failed. Otherwise, the nonzero coordinate is our estimator for the class associated with $\mathbf{x}$.

V. Face recognition using binary networks of tagged neural cliques

A. Storing (Learning) face images in binary networks of neural tagged cliques

Let us consider a set of training face images $\mathrm{S} = \{\mathrm{S}_i\}_{i=1}^{L}$ of cardinality L. We denote by $c \leq L$ the number of distinct persons (classes). For each person k, we compute the set F_k of SIFT features of all their corresponding images. We then index the set of all features $\mathrm{F} \triangleq \bigcup_{1\leq k\leq \mathrm{c}} \mathrm{F}_k = \{\mathbf{f}_1, \ldots, \mathbf{f}_d\}$.

We define the gallery vectors $\mathbf{g}_k$ by choosing $\mathbf{x}_k$ as the indicator vector of the subset F_k:

$$(\mathbf{x}_k)_i = \begin{cases} 1 & \text{if } \mathbf{f}_i \in \mathrm{F}_k \\ 0 & \text{otherwise} \end{cases} . \quad (9)$$

We afterwards perform the storage according to the method described in subsection IV-A. An example of a clique is shown in Figure 6. Such a graphical pattern is redundant and offer error correcting capabilities [8].

B. Retrieving face images in binary networks of neural tagged cliques

Let $\bar{\mathrm{S}} = \{\bar{\mathrm{S}}_i\}_{i=1}^{\bar{L}}$ be a set of face images to test. These images are novel but correspond to persons already seen in the gallery. First, each test face image $\bar{S}_i$ is described as a set of SIFT features $\bar{\mathrm{F}}_i$ We use the cosine similarity to compare two SIFT features as in (5). The input indicator vector $\mathbf{x}$ of the subset F is then defined as follows:

$$(\mathbf{x})_i = \begin{cases} 1 & \text{if } \exists\, \mathbf{f} \in \bar{\mathrm{F}}_i \text{ such that } i = \underset{j}{\operatorname{argmin}}\ \theta(\mathbf{f}, \mathbf{f}_j) \\ 0 & \text{otherwise} \end{cases} . \quad (10)$$

for i from 1 to d. According to the two-step Algorithm 1, the input pattern $\mathbf{x}$ is then retrieved and classified.

VI. Experimental evaluation

The Olivetti and Oracle Research Laboratory (ORL) face database is used in order to test our method in the presence of headpose variations. There are 10 different images of each of 40 distinct subjects. For some subjects, the images were taken at different times, varying lighting, facial expressions (open / closed eyes, smiling / not smiling), facial details (glasses / no glasses) and head pose (tilting and rotation up to 20 degrees). All the images were taken against a dark homogeneous background. Figure 5 shows the whole set of 40 individuals, 1 images per person from the ORL database.

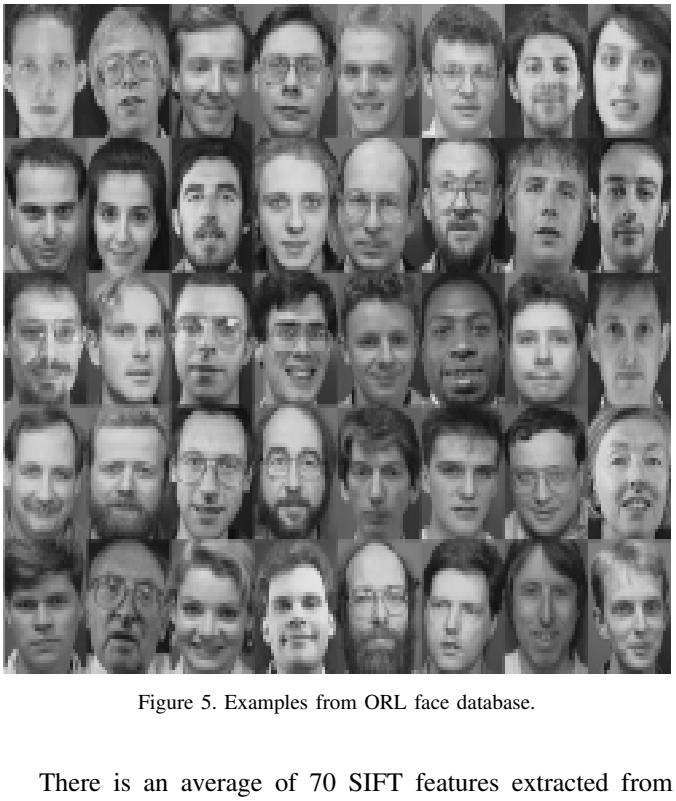


Figure 5. Examples from ORL face database.

There is an average of 70 SIFT features extracted from each image using the implementation proposed in [14]. Twenty independent runs were carried out. In each run, the dataset is randomly split into two halves, one for training (K images per class) and one for testing (the remaining $10 - K$ images per class).

Table I displays the results (average on the runs) obtained on the ORL database by several state of the art approaches, for comparison to the method proposed in this paper. All these methods are based on SIFT features. SIFT-based face recognition methods are actually more robust than other ones [10]. As shown in this table, the face recognition method introduced in this paper outperforms previously proposed approaches.

 ISBN: 978-1-61208-390-2

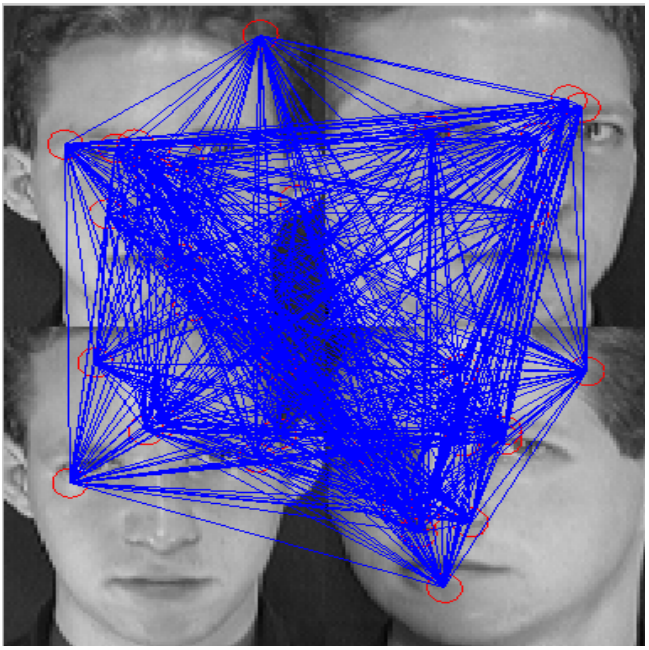


Figure 6. On this images, points of interests (represented as neurons in our model) have been fully interconnected to obtain a clique.

TABLE I. RECOGNITION RATES OF DIFFERENT MATCHING SCHEMES FOR THE ORL DATASET WITH RESPECT TO DIFFERENT SIZES FOR THE TRAINING SET.

Method	**Number of training images**			
	K = 5	K = 6	K = 7	K = 8
SIFT-SVM [12]	N/A	N/A	95.6	97.4
Aly [10]	92.42	95.27	96.88	98.36
Lenc-Kral [11]	96.75	97.86	98.65	98.86
Kepenekci [11]	97.92	97.86	98.65	99.17
Proposed method	**98.82**	**99.55**	**99.71**	**99.88**

It is worth pointing out that the decoding of a pattern in the clique network can be efficiently parallelized [15], [16].

VII. CONCLUSION AND FUTURE WORK

We successfully transformed the associative memory introduced in [9] into a classifier. Thanks to the error correcting properties of such memory, our proposed method outperforms state of the art SIFT-based face recognition approaches on the ORL database, without having to blow up the number of neurons. Since all the face recognition methods are based on the same feature descriptors, our results emphasize the interest of using clique-based networks as classifiers. It is worth pointing out that our method requires no threshold for face recognition, in contrast to the other ones.

Regarding scalability, future work involves assessing the impact of reducing network size towards performance. Compared to the theory of grandmother cells where a piece of information is carried out by a unique neuron, cliques offer the possibility to encode such data as an assembly of units. As a consequence the number of units needed to store information is significantly reduced in clique networks compared to perceptrons [17], for instance. Therefore one may expect complexity reduction compared to exhaustive search.

Regarding performance, we consider using complementary features to improve robustness of the recognition. In this respect, combining local SIFT descriptors with global features should increase accuracy of the system.

ACKNOWLEDGMENT

This work was partially funded as part of the NEUCOD project by the European Research Council under the European Unions Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement.

REFERENCES

[1] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, “Face recognition: A literature survey,” Acm Computing Surveys (CSUR), vol. 35, no. 4, 2003, pp. 399–458.

[2] M. Turk and A. Pentland, “Eigenfaces for recognition,” Journal of cognitive neuroscience, vol. 3, no. 1, 1991, pp. 71–86.

[3] P. N. Belhumeur, J. P. Hespanha, and D. Kriegman, “Eigenfaces vs. fisherfaces: Recognition using class specific linear projection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, 1997, pp. 711–720.

[4] B. Heisele, P. Ho, and T. Poggio, “Face recognition with support vector machines: Global versus component-based approach,” vol. 2. IEEE, 2001, pp. 688–694.

[5] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, “Robust face recognition via sparse representation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 2, 2009, pp. 210–227.

[6] L. Li, S. Li, and Y. Fu, “Learning low-rank and discriminative dictionary for image classification,” Image and Vision Computing, vol. 32, no. 10, 2014, pp. 814–823.

[7] D. Lowe, Object Recognition from Local Scale-Invariant Features. Springer Boston, 1999.

[8] V. Gripon and C. Berrou, “Sparse neural networks with large learning diversity,” IEEE Transactions on Neural Networks, vol. 22, no. 7, July 2011, pp. 1087–1096.

[9] B. K. Aliabadi, C. Berrou, V. Gripon, and X. Jiang, “Learning sparse messages in networks of neural cliques,” IEEE Transactions on Neural Networks and Learning Systems, August 2012.

[10] M. Aly, “Face recognition using sift features,” CNS/Bi/EE report, vol. 186, 2006.

[11] L. Lenc and P. Král, “Novel matching methods for automatic face recognition using sift,” Artificial Intelligence Applications and Innovations, 2012, pp. 254–263.

[12] L. Zhang, J. Chen, Y. Lu, and P. Wang, “Face recognition using scale invariant feature transform and support vector machine,” ICYCS’08: The 9th International Conference for Young Computer Scientists, 2008, pp. 1766–1770.

[13] J. J. Hopfield, “Neural networks and physical systems with emergent collective computational abilities,” Proceedings of the national academy of sciences, vol. 79, no. 8, 1982, pp. 2554–2558.

[14] A. Vedaldi, “An open implementation of the SIFT detector and descriptor,” no. 070012, 2007.

[15] H. Jarollahi, N. Onizawa, V. Gripon, and W. J. Gross, “Architecture and implementation of an associative memory using sparse clustered networks,” in IEEE International Symposium on Circuits and Systems (ISCAS), 2012, pp. 2901–2904.

[16] B. Larras, C. Lahuec, M. Arzel, and F. Seguin, “Analog implementation of encoded neural networks,” in IEEE International Symposium on Circuits and Systems (ISCAS), 2013, pp. 1612–1615.

[17] F. Rosenblatt, “The perceptron: a probabilistic model for information storage and organization in the brain.” Psychological review, vol. 65, no. 6, 1958, pp. 386–408.

 ISBN: 978-1-61208-390-2

Application of Loose Symmetry Bias to Multiple Meaning Environment

Ryuichi Matoba and Hiroki Sudo
Department of Electronics and Computer Engineering, National Institute of Technology, Toyama College
Email: {rmatoba, i08323}@nc-toyama.ac.jp

Makoto Nakamura
Japan Legal Information Institute, Graduate School of Law, Nagoya University
Email: mnakamur@law.nagoya-u.ac.jp

Satoshi Tojo
School of Information Science, JAIST
Email: tojo@jaist.ac.jp

***Abstract*—It is well known that the cognitive biases much accelerate the vocabulary learning. In addition, other works suggest that cognitive biases help to acquire grammar rules faster. The efficacy of the cognitive biases enables infants to connect an utterance to its meaning; even a single uttered situation contains many possible meanings. In this study, we focus on the symmetry bias which is one of the cognitive biases. The aim of this study is to evaluate the efficacy of the symmetry bias in the multiple meaning environment. In the experiments, two symmetry bias patters are utilized to evaluate the developed Meaning Selection Iterated Learning Model. The patterns are strict/loose symmetry bias with distance in languages and expressivity.**

***Keywords*—Symmetry Bias; Iterated Learning Model; Language Acquisition.**

I. INTRODUCTION

When learning a foreign language, learners might translate foreign words into their first language verbatim using a dictionary, and looking up grammar textbooks, or getting taught the foreign language by somebody who has already acquired it. In other words, learners use their first language to grasp meanings of the foreign language. In the case of the first language acquisition, learners do not know any languages to translate to figure out meanings of input utterances. From this aspect, the first language acquisition is more difficult than the second language, since an infant has no way to understand the meaning of each utterance. Thus, the infant has to identify the parent's intention from the situation. Under an environment which contains possibilities of many missteps to infer the parent's intention, it is hard to imagine that the infant smoothly acquires the first language.

In spite of this situation, infants can acquire new words very rapidly and also learn a word's meaning after just a single exposure [1], through fast mapping [2]. Under the above circumstance, various kinds of cognitive biases such as the shape bias [3] [4], the mutual exclusivity bias [5] [6], the whole object bias [7] and so on, work for infants to limit the possible word meanings [8] [9]. The definition of the cognitive bias is in the following quotation [10]:

> **Cognitive bias** Systematic error in judgment and decision–making common to all human beings which can be due to cognitive limitations, motivational factors, and/or adaptations to natural environments.

In this paper, we especially focus on the symmetry bias, and investigate its efficacy using computer simulation. The symmetry bias makes a strong relation between an object and its label by characterizing mapping among them as symmetric, i.e., it allows infants who are taught that a red sphere has a lexical label "apple", to make the reverse implication on their own, namely that the label "apple" refers to the red sphere object [11]. This tendency is said to be one of the peculiar human skills, and many experiments have endorsed that other animals cannot do this reverse implication [12].

This study aims to evaluate the efficacy of the symmetry bias using computer simulation. So far, we simulated the efficacy of the symmetry bias [13] using Iterated Learning Model (ILM) [14], and constructed joint attention frame in learning environment of infant agents [15]. Moreover, we formulated a method for measuring language distance to indicate the efficacy of the symmetry bias [16], and constructed the Meaning Selection Iterated Learning Model (MSILM) where a pair of a parent agent and an infant agent resides in a generation, and the infant agent becomes the parent agent of the next generation. In MSILM, the parent agent and the infant agent are given multiple meanings under the situation, where two agents share a common attention. The symmetry bias in our model works is based on the similarity of utterances.

In the last few years, our study suggests that the symmetry bias which connects a grammar and a meaning with complete symmetry does not accelerate effective grammar acquisition. In this paper, using MSILM, we evaluated the efficacy of two patterns of the symmetry bias which are strict/loose symmetry bias, which will be explained later.

This paper is organized as follows. In Section II, we introduce ILM and MSILM. In Section III, we examine our proposed method, and conclude in Section IV.

II. ILM WITH MEANING SELECTION

A. Briefing Kirby's ILM

Our study is based on ILM by Simon Kirby [14], who introduced the notions of compositionality and recursion as fundamental features of grammar, and showed that they made it possible for a human to acquire compositional language. Also, he adopted the idea of two different domains of language [17]–[19] which are I-language and E-language. The former is the internal language corresponding to the speaker's intention or meaning, while the other is the external language, that is, utterances. In Kirby's ILM, a speaker is a parent agent and a listener is an infant agent. The speaker agent gives the

 ISBN: 978-1-61208-390-2

listener agent a pair of a string of symbols as an utterance (E–language), and a predicate-argument structure (PAS) as its meaning (I–language). The agent's grammar is a set of a pair of a meaning and a string of symbols, as shown in formula (1).

$$S/love(john, mary) \rightarrow \texttt{lovejohnmary} \tag{1}$$

where the meaning, that is the speaker's intension, is represented by a PAS love(john, mary) and the string of symbols is the utterance "lovejohnmary"; the symbol 'S' stands for the category Sentence. The following rules can also generate the same utterance.

$$S/love(x, mary) \rightarrow \texttt{love}\,\texttt{N/x}\,\texttt{mary} \tag{2}$$

$$N/john \rightarrow \texttt{j} \tag{3}$$

where the variable x can be substituted for an arbitrary element of category N.

A number of utterances would form compositional grammar rules in a listener's mind, through the learning process. This process is iterated generation by generation, and finally, a certain generation would acquire a compact, limited number of grammar rules. The learner agent has the ability to generalize his/her grammar with learning. The learning algorithm consists of the following three operations; *chunk*, *merge*, and *replace* [14].

1) Chunk: This operation takes pairs of rules and looks for the most–specific generalization. For example,

$$\begin{cases} S/read(john, book) \rightarrow \texttt{ivnre} \\ S/read(mary, book) \rightarrow \texttt{ivnho} \end{cases} \tag{4}$$

$$\Downarrow$$

$$\begin{cases} S/read(x, book) \rightarrow \texttt{ivn}N/x \\ N/john \rightarrow \texttt{re} \\ N/mary \rightarrow \texttt{ho} \end{cases} \tag{5}$$

A rule without variables, i.e., the whole signal indicates the whole meaning of a sentence is called a *holistic rule*, while a rule with variables is called a *compositional rule*. In the case of the above example, two holistic rules become one compositional rule and two holistic rules by chunk operation.

2) Merge: If two rules have the same meanings and strings, replace their nonterminal symbols with one common symbol.

$$\begin{cases} S/read(x, book) \rightarrow \texttt{ivn}A/x \\ A/john \rightarrow \texttt{re} \\ A/mary \rightarrow \texttt{ho} \\ S/eat(x, apple) \rightarrow \texttt{apr}B/x \\ B/john \rightarrow \texttt{re} \\ B/pete \rightarrow \texttt{wqi} \end{cases} \tag{6}$$

$$\Downarrow$$

$$\begin{cases} S/read(x, book) \rightarrow \texttt{ivn}A/x \\ A/john \rightarrow \texttt{re} \\ A/mary \rightarrow \texttt{ho} \\ S/eat(x, apple) \rightarrow \texttt{apr}A/x \\ A/pete \rightarrow \texttt{wqi} \end{cases} \tag{7}$$

3) Replace: If a rule can be embedded in another rule, replace the terminal substrings with a compositional rule.

$$\begin{cases} S/read(pete, book) \rightarrow \texttt{ivnwqi} \\ B/pete \rightarrow \texttt{wqi} \end{cases} \tag{8}$$

$$\Downarrow$$

$$\begin{cases} S/read(x, book) \rightarrow \texttt{ivn}B/x \\ B/pete \rightarrow \texttt{wqi} \end{cases} \tag{9}$$

In Kirby's experiment [14], five predicates and five object words are employed. Also, two identical arguments in a predicate like "hate(mary, mary)" are prohibited. Thus, there are 100 distinct meanings (5 predicates $\times$ 5 possible first arguments $\times$ 4 possible second arguments) in a meaning space.

Since the number of utterances is limited to 50 in his experiment, the infant agent cannot learn the whole meaning space, the size of which is 100; thus, to obtain the whole meaning space, the infant agent has to generalize his/her own knowledge by self-learning, i.e., chunk, merge, and replace. The parent agent receives a meaning selected from the meaning space, and utters it using her own grammar rules. When the parent agent cannot utter because of lack of her grammar rules, she invents a new rule. This process is called invention. Even if the invention does not work to complement the parent agent's grammar rules to utter, she utters a randomly composed sentence.

B. Briefing MSILM

Our model, MSILM (see Figure 1), introduces the notion of joint attention frame, as mentioned the previous section, into the ILM. In MSILM, multiple meanings are presented to both the parent and the infant agent, and the parent agent mentions one of them. The infant agent listens to the utterance from the parent agent, and infers its meaning from the presented meanings using an inference strategy, i.e., the symmetry bias. This model represents a situation in which the infant agent does not always acquire a unique meaning of a parent's utterance.

In our model, we changed two points of Kirby's model, which are (i) taking away the transmittance of meanings between the parent and the infant, and (ii) introducing a set of multiple meanings which contains more than one meaning. This would cause a significant difference from the result of ILM, namely the infant agent has a possibility to connect a parent's utterance to a meaning which is not that of the parent's intention of the utterance, and this leads the infant to acquire a far different grammar from the parent.

So far, following the evaluation method of Kirby, we have only used expressivity which is defined as the ratio of the number of utterable meanings derived from the grammar rules to the whole meaning space, and the number of rules of grammar to evaluate agent's grammar efficacy. However, our motivation of this study is to evaluate the efficacy of the symmetry bias by measuring the differences of grammar between the infant and the parent in a quantitative way. Therefore, we have introduced the distance in languages as well as expressivity as an evaluating method for the infant's acquired grammar. For evaluating the distance of two grammars, we define the distance in languages by the edit distance, known as the Levenshtein distance; we count the number of insertion/elimination operations to change one word into the

 ISBN: 978-1-61208-390-2

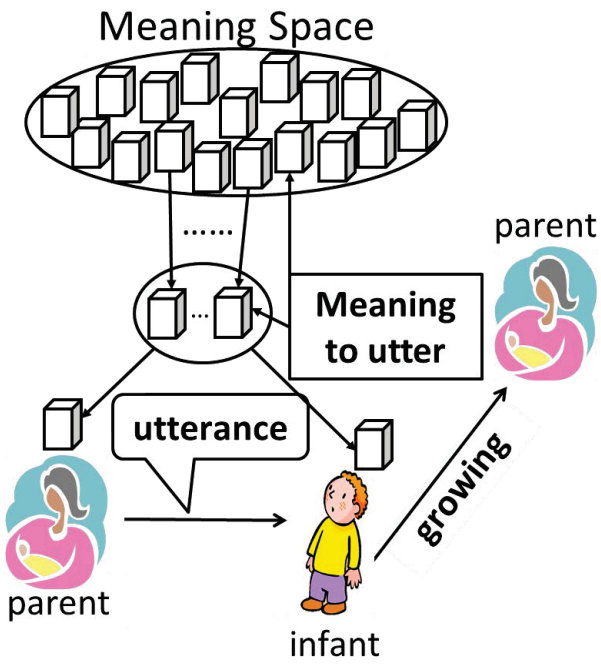


Fig. 1. Illustration of Meaning Selection Iterated Learning Model.

another. For example, the distance between 'abc' and 'bcd' becomes 2 (erase 'a' and insert 'd'). All the compositional grammar rules are expanded into a set of holistic rules, which do not include any variable, i.e., a rule consists of a sequence of terminal symbols. Now the comparison between a parent agent and an infant agent takes the following procedure.

1) Pick up a grammar rule (g_c) which is constructed by a pair of a PAS (p_c) and an utterance (u_c) from the child's grammar rules (G_c). Choose a grammar rule ($g_p^{p_c}$) in which PAS ($p_p^{p_c}$) is the most similar to p_c from parent's grammar rules (G_p), in terms of the Levenstein distance. If there are multiple candidates, all of them are kept for the next process.
2) Focus on an utterance ($u_p^{p_c}$) of $g_p^{p_c}$ and u_c, and measure a distance ($d(u_c, u_p^{p_c})$ between $u_p^{p_c}$ and u_c using the Levenshtein distance. If there are multiple candidates, choose the smallest one.
3) Normalize d from 0 to 1.
4) Carry out 1 to 3 for all grammar rules of G_c. Calculate the sum of all the distances and regard the average of them as the distance of two sets of linguistic knowledge. Thus, in this case, the distance between G_c and G_p is calculated as Formula (10).

$$Dist_{G_c to G_p} = \frac{1}{|G_c|}\left(\sum_{i=0}^{|G_c|}\frac{d(u_{ci}, u_p^{p_{ci}})}{|u_{ci}| + |u_p^{p_{ci}}|}\right) \qquad (10)$$

The image of this measuring procedure is shown in Figure 2.

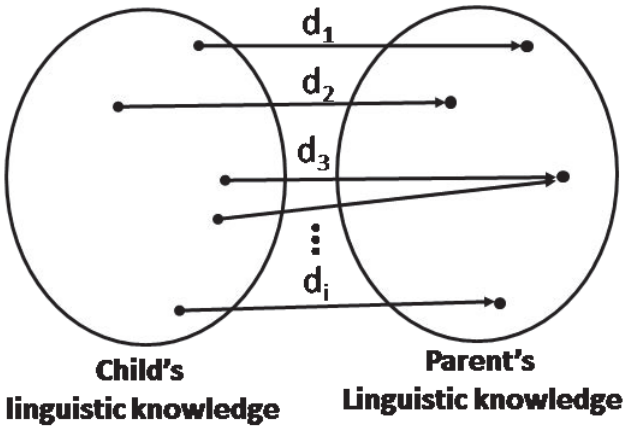


Fig. 2. Image of measuring procedure.

III. Experiment and Result in MSILM

In the simulation, preserving Kirby's settings, we employed five predicates and five object words, also prohibited two identical arguments in a predicate. This implies the size of meaning space of the experimental model is 100. A point of difference is the number of meanings which are presented to the agents. In the experiment, two meanings are presented to both the parent and the infant agent, and we examined the following three strategies when the infant agent infers the meaning of a parent's utterance from two meanings.

1) **Random**: The infant agent chooses a meaning from presented meanings randomly as a meaning of a parent's utterance.
2) **Strict Symmetry Bias**: If the infant agent can generate the same utterance as the utterance from the parent agent using own grammar, and its meaning is found in the presented meanings, he/she connects the utterance and its meaning. Otherwise, the infant agent employs the random strategy.
3) **Loose Symmetry Bias**: If strict strategy fails, the infant agent compares all utterances which he/she can generate to the parent's utterance using Levenshtein distance, and chooses the most similar one. Next, he/she compares a meaning of the selected utterance to presented meanings, and chooses the most similar meaning from the presented meanings.

Figures 3 and 4 show the average tendency of expressivity and the distance in languages per each generation, after 100 trials. Each line denotes the result of the strategies of random, strict symmetry bias and loose symmetry bias, respectively.

From Figure 3, we can observe that expressivity of the agent who takes the loose strategy records the highest value of the three strategies, also, the strict strategy is the lowest despite the infant agent applies the symmetry bias. In the case of applying strict symmetry bias, the infant agent receives information that he/she already knows from the parent agent, i.e., he/she does not get new information. Therefore, expressivity of the infant agent who employs strict strategy records the lowest.

 ISBN: 978-1-61208-390-2

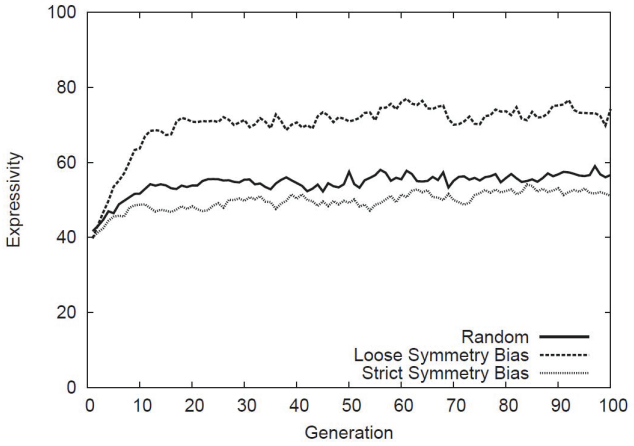


Fig. 3. The movement of the expressivity per generation.

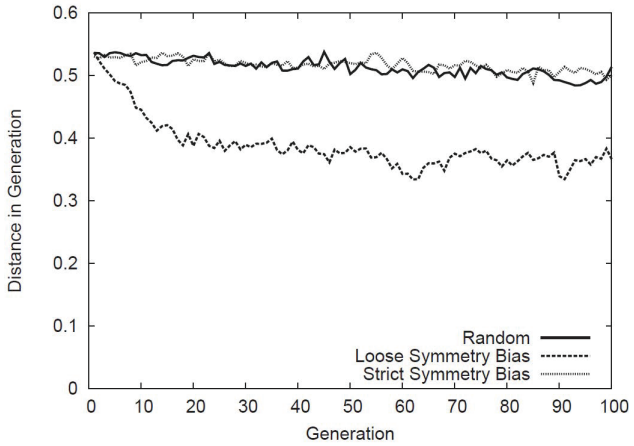


Fig. 4. The movement of the expressivity per generation.

From Figure 4, we can observe that the distance of loose strategy is the smallest, i.e., grammar of the infant agent and the parent agent is the most similar of the three strategies. For the above reasons, loose symmetry bias is the most effective strategy to acquire the parent's grammar of the three strategies.

IV. Conclusion

In this study, we verified the efficacy of the symmetry bias not only in the lexical acquisition, but also in the grammar acquisition. For this purpose, we have revised Kirby's model [14] to MSILM, and have built two kinds of symmetry bias, which are strict symmetry bias and loose symmetry bias. In the simulation, both of the parent and the infant agent are presented with multiple meanings, and the parent agent chooses one of them to utter. The infant agent receives an utterance and infers its meaning with three kinds of strategies which are random, strict symmetry bias, and loose symmetry bias.

For each of the strategies, we have observed expressivity, and the distance in languages. As a result of the experiments, the infant agent who has employed loose strategy,

- could acquire the highest expressivity,
- could construct the most similar grammar to his/her parents.

Our future works are summarized as follows. So far, we have only implemented the symmetry bias to a computer simulation model, and not compared it to a phenomenon of actual world yet. We should describe the efficacy of the symmetry bias of our model based in a real world experience.

Acknowledgment

This work was partly supported by Grant-in-Aid for Young Scientists(B)(KAKENHI)No. 23700310, and Grant-in-Aid for Scientific Research(C)(KAKENHI)No.25330434 from MEXT Japan.

References

[1] S. Carey and E. Bartlett, "Acquiring a single new word," Child Language Development, vol. 15, 1978, pp. 17–29.

[2] J. S. Horst and L. K. Samuelson, "Fast mapping but poor retention by 24-month-old infants," Infancy, vol. 13, 2008, pp. 128–157.

[3] B. Landau, L. B. Smith, and S. S. Jones, "The importance of shape in early lexical learning," Cognitive Development, vol. 3, no. 3, 1988, pp. 299–321.

[4] ——, "Syntactic context and the shape bias in children's and adult's lexical learning," Journal of Memory and Language, vol. 31, no. 6, 1992, pp. 807–825.

[5] E. M. Markman, "Constraints children place on word meanings," Cognitive Science: A Multidisciplinary Journal, vol. 14, no. 1, 1990, pp. 57–77.

[6] E. M. Markman, J. L. Wasow, and M. B. Hansen, "Use of the mutual exclusivity assumption by young word learners," Cognitive Psychology, vol. 47, no. 3, 2003, pp. 241–275.

[7] E. M. Markman, Categorization and naming in children: Problems of induction. Cambridge: MIT Press, 1989.

[8] M. Imai and D. Gentner, "Children's theory of word meanings: The role of shape similarity in early acquisition," Cognitive Development, vol. 9, no. 1, 1994, pp. 45–75.

[9] ——, "A crosslinguistic study of early word meaning: Universal ontology and linguistic influence," Cognition, vol. 62, no. 2, 1997, pp. 169–200.

[10] A. Wilke and R. Mata, Cognitive Bias. Academic Press, 2012, vol. 1.

[11] M. Sidman, R. Rauzin, R. Lazar, S. Cunningham, W. Tailby, and P. Carrigan, "A search for symmetry in the conditional discriminations of rhesus monkeys, baboons, and children," Journal of the Experimental Analysis of Behavior, vol. 37, 1982, pp. 23–44.

[12] Y. Yamazaki, "Logical and illogical behavior in animals," Japanese Psychological Research, vol. 46, no. 3, 2004, pp. 195–206.

[13] R. Matoba, M. Nakamura, and S. Tojo, "Efficiency of the symmetry bias in grammar acquisition," Information and Computation, vol. 209, 2011, pp. 536–547.

[14] S. Kirby, Learning, Bottlenecks and the Evolution of Recursive Syntax. Cambridge University Press, 2002.

[15] R. Matoba, S. Shoki, and H. Takashi, "Cultural Evolution of Compositional Language under Multiple Cognition of Meanings," in Proceedings of the 15th International Symposium on Artificial Life and Robotics(AROB 2010), 2010.

[16] R. Matoba, H. Sudo, S. Hagiwara, and S. Tojo, "Evaluation of Efficiency of the Symmetry Bias in Grammar Acquisition," in Proceedings of the 18th International Symposium on Artificial Life and Robotics(AROB 2013), 2013, pp. 444–447.

[17] J. Hurford, Ed., Language and Number: the Emergence of a Cognitive System. Blackwell, 1987.

[18] D. Bickerton, Ed., Language and Species. University of Chicago Press, 1990.

[19] S. Kirby, Function, Selection, and Innateness: The Emergence of Language Universals. Oxford University Press, 1999.

 ISBN: 978-1-61208-390-2

An Experimental Investigation on Learning Activities Inhibition Hypothesis in Cognitive Disuse Atrophy

Kazuhisa Miwa
Graduate School of Information Science
Nagoya University
Nagoya, JAPAN
miwa@is.nagoya-u.ac.jp

Kazuaki Kojima
Learning Technology Laboratory
Teikyo University
Utsunomiya, JAPAN
kojima@lt-lab.teikyo-u.ac.jp

Hitoshi Terai
Graduate School of Information Science
Nagoya University
Nagoya, JAPAN
terai@is.nagoya-u.ac.jp

***Abstract*—The term "disuse atrophy" is generally used for physical atrophy such as muscle wasting. When muscles are no longer used, they slowly weaken. This weakening, or atrophy, can also occur from continuous physical support that leads to a minimal use of the body. We advance the idea that disuse atrophy occurs not only in the physical realm but also in cognitive ability. We investigate why cognitive disuse atrophy occurs. Specifically, we examine the learning activities inhibition hypothesis, which posits that cognitive disuse atrophy occurs because continuous use of support systems provides cognitive shortcuts for performing activities and inhibits learning-oriented activities. To investigate this hypothesis, two experiments were performed in which the participants played Reversi games. Both Experiments 1 and 2 indicated that the participants' winning rates were highest when they were given a higher level of support, and their decision times for determining each move were shortest in the training phase. Experiment 2 also indicated that participants' post-test scores (measured as learning gains) were lower when they were given higher levels of support. These results confirmed that a higher level of support promotes performance-oriented activities, but inhibits learning-oriented activities when engaging in training, supporting the learning activities inhibition hypothesis.**

Keywords–cognitive disuse atrophy; performance-orienged activities; learning-orienged activities

I. INTRODUCTION

A variety of human support systems based on advanced technologies, such as automation systems, operate in our daily lives. These systems have contributed to the increase of human abilities to perform tasks. However, we often recognize negative secondary effects of the overuse of such systems (e.g., difficulty in memorizing maps due to daily usage of car navigation systems, or difficulty remembering the accurate spelling of words because of using a word processor with spell checker software). Human factor studies have reported that the continuous use of automated systems decreases users' manipulation abilities [1][2], and more seriously, complacency on this front causes aircraft accidents [3]. This happens because long-term continuous supports decrease human cognitive activity, weakening the ability to performing tasks.

Miwa and Terai proposed the concept of cognitive ability disuse atrophy, the loss of cognitive ability due to the disuse of cognitive activities [4]. We see this as a key issue underlying some human factor problems that emerge when people engage in cognitive tasks aided by computers. The term "disuse atrophy" is generally used for physical atrophy, such as muscle wasting [5]. We advance the idea that disuse atrophy occurs not only in the physical realm but also in cognitive ability.

In this paper, we investigate why cognitive disuse atrophy occurs. Specifically, we propose the learning activities inhibition hypothesis to explain this psychological phenomenon. In explaining this hypothesis, we first note the duality of cognitive processing when engaging in a task [6]. Generally, there are two objectives for performing a task. One ordinary objective is to perform and complete the task. However, there is another important objective: for performers to develop proficiency and knowledge by performing the task. Performance and mastery are the prime reasons to engage in a task. We contend that cognitive disuse atrophy emerges when the mastery factor is lost.

For example, consider car navigation systems. When people search for a route from a current location to a new destination, they usually try to remember a mental map, a configuration of the possible pathways, select candidate pathways related to the target route, and decide on the best route from multiple candidates while considering current traffic and construction. These cognitive information processing efforts develop a mastery of memorizing maps and the acquisition of the skills to search for a route. However, when we use navigation systems, we do not need to perform any such mental activities. All one has to do is to enter the destination and press the confirmation button. From the perspective of performance, this is all it takes to achieve the goal. But for mastery, the mental activities of memorizing a map and finding a route with a printed map are also important. Since car navigation systems deprive users of opportunities for such efforts toward mastery, they cause mental disuse atrophy.

We define performance-oriented activities as those for performing tasks and learning-oriented activities as those for mastery. The learning activities inhibition hypothesis proposes that cognitive disuse atrophy occurs because the continuous use of support systems provides cognitive shortcuts for performing activities and thus inhibits the learning-oriented activities.

In this paper, we empirically investigate the learning activities inhibition hypothesis in the following research paradigm. We had participants engage in a task. In the training phase, participants performed the task with help from a task-supporting system. Task performance was measured and used as an index

 ISBN: 978-1-61208-390-2

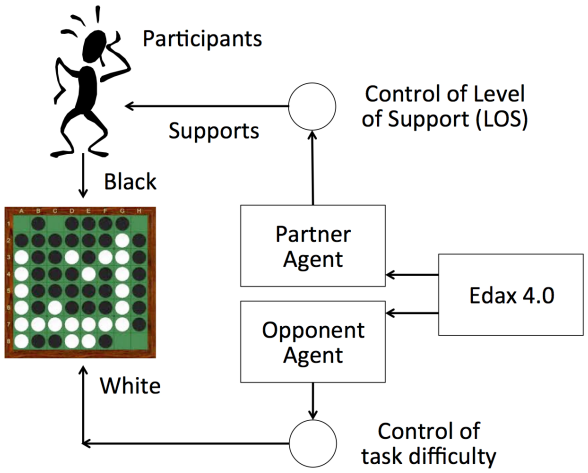


Figure 1. Overall configuration of the Reversi-based learning environment.

for their performance-oriented activities. After the training phase, a post-test was performed without any supports available. Post-test scores were measured and used as an index of their learning-oriented activities in the training phase. We then evaluated the two indexes as a function of the level of support (LOS) in the training phase.

The learning activities inhibition hypothesis predicts the following:

- As LOS increases, task performance in the training phase would increase because the performance-oriented activities would increase due to high-level supports.
- However, post-test scores would decrease because high-level assistance inhibits the learning-oriented activities in the training phase.

In Section 2, we explain an experimental system developed for this study, and Experiment 1 in Sections 3 and 4 and Experiment 2 in Sections 5 and 6 are reported, followed by discussion and conclusions in Section 7.

II. EXPERIMENTAL SYSTEM

A. *Reversi-based learning environment*

We developed a Reversi-based learning environment as a workbench to investigate the learning activities inhibition hypothesis. Figure 1 shows the overall configuration of the experimental system. In our experimental environment, a participant plays 8 by 8 Reversi games against a virtual opponent (i.e., opponent agent) on a computer. A virtual partner (i.e., partner agent) assists the participant in selecting winning moves. Both agents, opponent and partner, are controlled by a Reversi engine, Edax, which suggests the best moves by assessing future states in the game. The opponent's competence can be controlled by setting the maximum depth to which Edax searches for future game states. The partner agent recommends candidate moves among valid squares before the participant makes a move.

The Edax-generated opponents are exceptionally competent Reversi players that cannot be defeated by human participants.

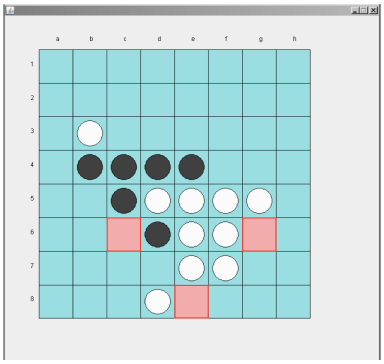


Figure 2. An example screenshot of the experimental system.

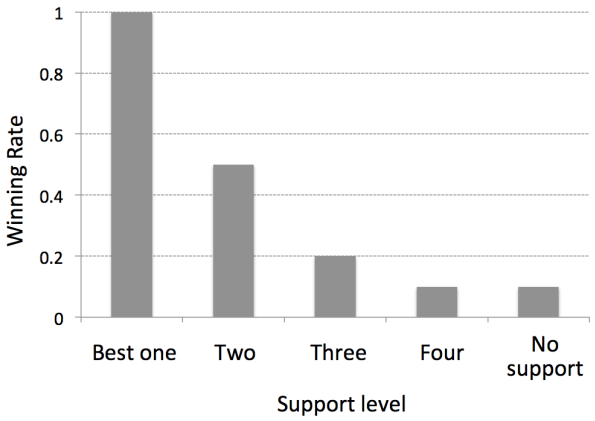


Figure 3. Winning rate of the simulated participants against an opponent agent.

To reduce the strength of the opponent agent to a level compatible with human participants, the agents were set to randomly miss the best move twice in the initial and middle stages. Support levels (LOSs) from the partner agent were controlled by presenting the candidate with the best or multiple moves, or no candidates (i.e., no supports). Figure 2 shows an example screenshot of the experimental system where three candidate moves are presented.

B. *Preliminary simulations*

To predict the degree of winning by human participants in the environment, we conducted preliminary simulations. The simulated participant randomly selected one of the candidate moves. In the no-support condition, it randomly selected one of the possible moves.

Figure 3 shows the ratio of wins by the simulated participant against the opponent agent in 20 simulated games. This figure implies that the winning ratio of human novices increases as the support level increases. However, the learning activities inhibition hypothesis predicts that consistently presenting the best move to participants would inhibit their skill mastery. Therefore, post-test scores in the best-move presentation condition would be lower than those in the multiple-candidate-moves presentation condition and the no-supports condition.

 ISBN: 978-1-61208-390-2

III. EXPERIMENT 1

A. Participants

A total of 71 undergraduate students in the school of informatics and sciences at Nagoya University participated in Experiment 1. They were paid 4000 Japanese Yen as baseline, and were additionally paid to a maximum of 3000 Yen based on their performance measured as post-test scores.

B. Experimental conditions

We manipulated the LOS in participants' training by setting up three experimental conditions: (1) the Best Move condition, where the partner agent suggested the best move to the participants, (2) the Three Candidates condition, where three candidate moves were suggested, including the best move, and (3) the No Support condition, where no suggestions were given.

Twenty-three, twenty-four, and twenty-four participants were assigned to the Best Move, Three Candidates, and No Support conditions, respectively.

C. Experimental procedure

In the initial stage, participants were instructed on how to operate the experimental system. In the Best Move and Three Candidates conditions, participants were taught that a virtual partner would present candidate moves in each trial, but they are not required to follow the suggestions. After the instruction phase, a pre-test was performed. The participants played a game against the virtual opponent without the partner agent's supports.

In the training phase, participants were divided into three groups and played twelve games in which the LOS was controlled. After the training phase, a post-test was performed in which the experimental setting was identical to that of the pre-test phase. After the post-test was performed, the participants played four additional games in each of the experimental settings, and then performed the second round of post-test. We evaluated their learning gains based on the first and second rounds of post-tests.

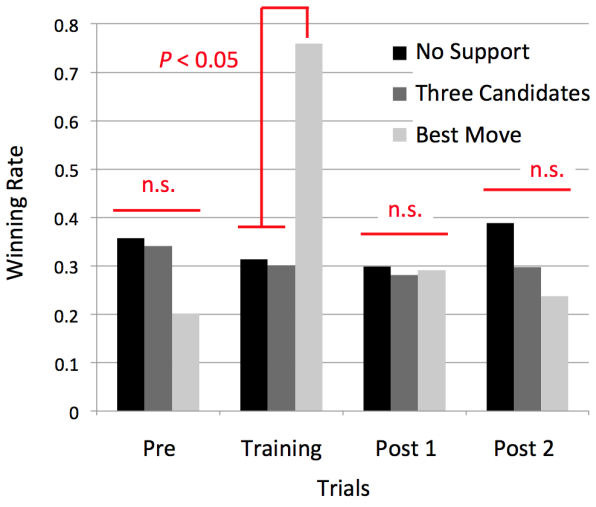


Figure 4. Winning rate as a function of experimental conditions.

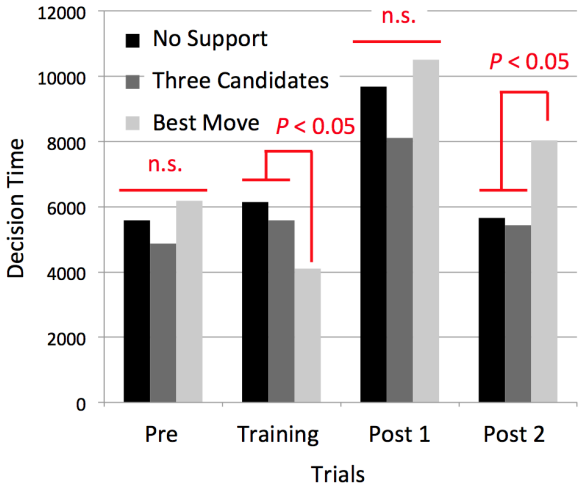


Figure 5. Decision time as a function of experimental conditions.

IV. RESULTS

A. Winning Rate

First, we evaluated the participants' performance in the training phase. Figure 4 shows the winning rate (the ratio of the obtained pieces (black pieces) to the total number of pieces (i.e., black and white pieces)) in the pre-test, the training phase, and the first and second rounds of post-tests.

A three (Condition: No Support, Three Candidates, and Best Move) × four (Trials: Pre, Training, Post 1, and Post 2) ANOVA revealed a significant interaction ($F(6, 204) = 13.11$, $p < 0.01$). The simple main effect of the Condition factor did not reach a significant level at Pre, Post 1, and Post 2 ($F(2, 68) = 3.12$, n.s.; $F(2, 68) < 1$, n.s.; $F(2, 68) = 1.94$, n.s.), but revealed significance at Training ($F(2, 68) = 184.74$, $p < 0.01$). The LSD analysis indicated that the winning rate was significantly higher in the Best Move condition than those in the Three Candidates and No Support conditions ($p < 0.05$; $p < 0.05$).

B. Decision Time

Second, we evaluated the participants' behavior based on the average time for determining one move in their turn. First, we calculated the average time to decide one move in a game; then we averaged the decision times over the twelve games. Figure 5 shows the result.

Again, a three (Condition: No Support, Three Candidates, and Best Move) × four (Trials: Pre, Training, Post 1, and Post 2) ANOVA revealed a significant interaction ($F(6, 204) = 4.36$, $p < 0.01$). The simple main effect of the Condition factor did not reach a significant level at Pre and Post 1 ($F(2, 68) = 1.63$, n.s.; $F(2, 68) = 2.27$, n.s.), but revealed significance at Training and Post 2 ($F(2, 68) = 5.13$, $p < 0.01$; $F(2, 68) = 3.44$, $p < 0.05$). At Training, the LSD analysis indicated that the decision times in the No Support and Three Candidates conditions were longer than that in the Best Move condition ($p < 0.05$; $p < 0.05$). In contrast, at Post 2, the decision times in

 ISBN: 978-1-61208-390-2

the No Support and Three Candidates conditions were shorter than that in the Best Move condition ($p < 0.05$; $p < 0.05$).

C. Discussion

The learning activities inhibition hypothesis predicted that the winning rate in the training phase would increase with higher support conditions. This prediction was partially confirmed because the rate was highest in the Best Move condition, but no difference was found between the No Support and Three Candidates conditions.

The hypothesis also predicted that post-test scores would be higher in lower support conditions; but this prediction was not confirmed. There were no significant differences in the winning rates in the post-test among the three conditions. However, note that for Post 2, decision times were longer in the Best Move condition than in the other two lower support conditions. This implies that training with such a high level of support, where participants were continuously given the best move, may have inhibited learning gains, resulting in longer decision times in the post-test phase where no such computer supports were available. This speculation is consistent with the shorter decision times for Training in the Best Move condition, implying that shorter decision times reflect superficial thinking without deliberate consideration during training.

V. Experiment 2

The overall results in Experiment 1 confirmed that the performance-oriented activities are raised in higher-supported situations, but not that the learning-oriented activities increase in less-supported situations. In terms of learning gains, shorter decision times in lower supported conditions were found only in Post 2 after additional four training trials, but not in Post 1. This may imply that learning effects may emerge after longer training times. Based on this insight, we conducted Experiment 2.

A. Participants

Initially, 27 undergraduate students in the school of informatics and sciences at Nagoya University participated in Experiment 2. They were not paid because the experiment was performed as a part of the class curricula for cognitive science. Twenty-one participants were analyzed since six of the initial participants withdrew from the experiment after the pre-test.

B. Experimental conditions

In Experiment 1, we could not confirm any differences between the Three Candidates and No Support Conditions. Therefore, in Experiment 2, we set up only two experimental conditions: the Best Move and No Support conditions. The initial 27 participants were ordered according to their pre-test scores and were divided into two groups. Specifically, odd-numbered (i. e., top, third, fifth, etc.) participants were assigned to one of the two conditions, and even-numbered participants were assigned to the other condition. Six participants withdrew from the experiment, resulting in nine and twelve participants working in the No Support and Three Candidate conditions, respectively.

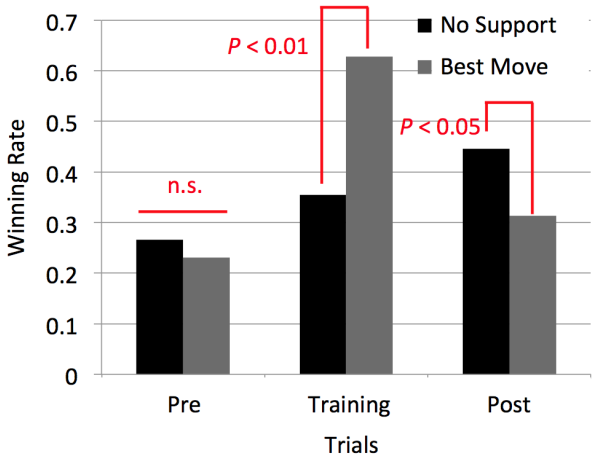


Figure 6. Winning rate as a function of experimental conditions.

C. Experimental procedure

In the initial stage, participants were instructed on how to operate the experimental system. In the Best Move condition, they were taught that a virtual agent would present the best move in each trial, but they were not required to follow its suggestions. After the instruction phase, a pre-test was given. Participants then played three games against the virtual opponent without the partner agents' supports.

After the pre-test, the participants were instructed to play three games a day over two weeks to train themselves. They were required to report their trials daily via e-mail to the experimenter. To ensure that the participants continuously engaged in games throughout two weeks, the experimenter sent e-mail reminders to participants if their daily e-mail reports were not received.

After two weeks, a post-test was performed. The experimental setting was identical to that of the pre-test phase. The participants played three games as a post-test.

VI. Results

A. Winning Rate

As in Experiment 1, we first evaluated participants' performance in the training phase. Figure 6 shows the winning rates in the pre-test, two-week training period, and the post-test.

In the following analysis, the average scores over the three games were used as the pre- and post-test scores. To calculate the training scores, we first calculated the average values of each day's three games, then averaged the values over two weeks.

A two (Condition: No Support and Best Move) $\times$ three (Trials: Pre, Training, Post) ANOVA revealed a significant interaction ($F(2, 38) = 8.59$, $p < 0.01$). The simple main effect of the Condition factor did not reach a significant level at Pre ($F(1, 19) < 1$, n.s.), but revealed significance at Training and Post ($F(1, 19) = 9.52$, $p < 0.01$; $F(1, 19) = 4.81$, $p < 0.05$), indicating that the winning rates during training were higher, but the rates in the post-test were lower in the Best Move condition than those in the No Support condition.

 ISBN: 978-1-61208-390-2

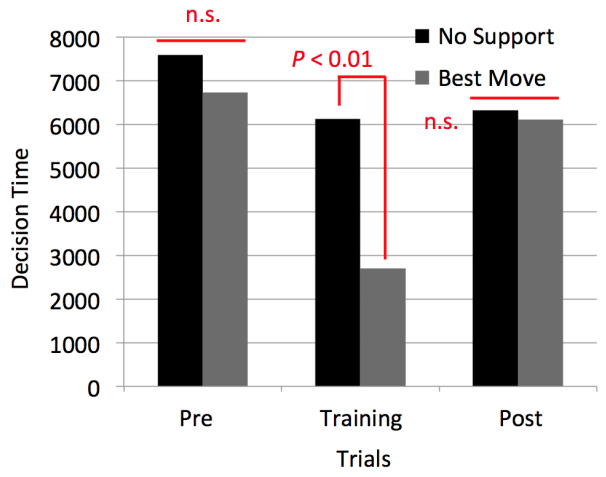


Figure 7. Decision time as a function of experimental conditions.

B. Decision Time

Second, we also evaluated the participants' behavior based on their average times for determining one move during their turn. Figure 7 shows the result.

A two (Condition: No Support and Best Move) × three (Trials: Pre, Training, and Post) ANOVA did not reveal a significant interaction ($F(2, 38) = 3.13$, n.s.). The main effect of the Condition factor did not reach a significant level ($F(1, 19) = 2.48$, n.s.). However, the figure obviously predicts a difference between the two experimental conditions in the training phase. Therefore, we performed individual statistical analyses at Pre, Training, and Post, respectively. The results show that the decision times at Training in the No Support condition were longer than that in the Best Move condition ($F(1, 19) = 15.82$, $p < 0.01$), even though there were no significant differences in decision times in the Pre and Post phases ($F(1, 19) < 1$, n.s.; $F(1, 19) < 1$, n.s.).

C. Discussion

In the training phase, the winning rates were higher and the decision times were shorter in the Best Move condition, confirming that the performance-oriented activities increase in a higher-support situation. More importantly, for the post-test, the winning rates in the Best Move condition were lower than that in the No Support condition, implying that the learning-oriented activities are inhibited in the Best Move condition. These results support the learning activities inhibition hypothesis.

VII. Discussion and Conclusions

A. Summary

In the training phase, Experiments 1 and 2 indicated that the participants' winning rates were the highest in the Best Move condition, and their decision times for determining each move were the shortest. Experiment 2 indicated that, the participants' post-test scores measured as learning gains were lower in the Best Move condition than that in the No Support condition. These results confirmed that a higher level of support promotes the performance-oriented activities, but inhibits the learning-oriented activities of training, thus supporting the learning activities inhibition hypothesis.

B. Assistance Dilemma

Similar findings have been reported in studies on intelligent tutoring systems. Koedinger and Aleven (2007) posed a crucial question: How should learning environments balance assistance and the withholding of assistance to optimize the learning process? [7] This assistance dilemma is considered a central topic for establishing instructional principles in tutoring. While high assistance provides useful scaffolding that sometimes facilitates problem solving in the learning phase, it also elicits superficial responses given without serious consideration. On the other hand, low assistance encourages self-learning in students, but may introduce major errors and sometimes impede problem solving in learning.

The assistance dilemma implies that, in some cases, reducing support levels increases learning effects, even while incurring a partial loss of problem-solving performance. This speculation is consistent with the findings confirmed in this paper.

C. Cognitive Load Theory

Cognitive load theory gives us another informative perspective about the cognitive mechanisms underlying the tradeoff between the performance-oriented and learning-oriented activities confirmed in this study.

Cognitive load theory has provided design principles for learning environments constrained by cognitive architecture. The theory distinguishes three types of cognitive loads: intrinsic, extraneous, and germane [8][9]. The intrinsic load is the basic cognitive load required to perform a task. The intrinsic load increases with the increasing difficulty of a task and the decreasing expertise of the performer. The extraneous load is the wasted cognitive load unrelated to learning activities, and is reluctantly processed. One source of extraneous load is inappropriately designed learning material. The extraneous load can also be increased by a lack of related knowledge and problem solving skills. Finally, the germane load is the cognitive load for learning, such as constructing schemata.

From the perspective of cognitive load theory, the intrinsic load presumably contributes to the performance-oriented activities, and the germane load contributes to the learning-oriented activities.

Adequate assistance decreases the extraneous load by presenting related information for problem solving. Many design principles for reducing the extraneous load have been proposed [8]. However, note that decreasing the extraneous load by providing high-level assistance does not necessarily increase the germane load when superficial problem solving without deliberate thinking is performed.

Producing the germane load sufficient for maximizing learning gains is a challenging problem [10][11]. Only a limited number of design principles exist for raising the germane load while maintaining sufficient intrinsic load for performing a task. In this study, we have presented a case in which low levels of assistance may guide students toward deeper consideration, activating their learning-oriented activities for expertise.

 ISBN: 978-1-61208-390-2

D. Future Work

To maximize learning gains, the balance between the performance-oriented and learning-oriented activities is crucial. It is important to find ways to manipulate the balance between these two types of cognitive activities. One important factor for such manipulation comes from goals established by performers.

Goal achievement theory has provided theoretical perspectives on the relationships between students' goals and their learning activities, and also accumulated a vast amount of empirical findings [12]. In goal achievement theory, students' goals are divided into mastery and performance goals. The former motivates students to develop their own abilities, while the latter motivates them to seek higher social evaluation rather than their own development. This implies that mastery goals activate the learning-oriented activities and performance goals activate the performance-oriented activities. In the early stages of goal achievement theory, mastery goals were found to be more important than performance goals [13][14].

A recent study found the relationship between students' goals and their seeking of computer support [15]. According to this study, mastery-goal-oriented students tend to seek abstract (i.e., low-level) supports first, and then move to more specific (high-level) supports, whereas performance-goal-oriented students preferred quick and direct supports during the initial stage. This implies that mastery oriented-students tend to focus their cognitive load on learning-oriented activities.

In our future work, it will be important to conduct further experiments controlling for participants' goal factors in order to seek a way to promote the learning-oriented activities while cultivating sufficient performance-oriented activities.

ACKNOWLEDGMENT

This research was partially supported by HAYAO NAKAYAMA Foundation for Science & Technology and Culture, and JSPS KAKENHI Grant Number 25560110.

REFERENCES

[1] R. Parasuraman and V. Riley, "Humans and automation: Use, misuse, disuse, abuse," Human Factors: The Journal of the Human Factors and Ergonomics Society, vol. 39, no. 2, 1997, pp. 230–253.

[2] N. B. Sarter and D. D. Woods, "Team play with a powerful and independent agent: a full-mission simulation study," Human Factors: The Journal of the Human Factors and Ergonomics Society, vol. 42, no. 3, 2000, pp. 390–402.

[3] E. L. Wiener and R. E. Curry, "Flight-deck automation: Promises and problems," Ergonomics, vol. 23, no. 10, 1980, pp. 995–1011.

[4] K. Miwa and H. Terai, "Theoretical investigation on disuse atrophy resulting from computer support for cognitive tasks," in Engineering Psychology and Cognitive Ergonomics, ser. Lecture Notes in Computer Science, D. Harris, Ed. Springer International Publishing, 2014, vol. 8532, pp. 244–254.

[5] S. W. Jones, R. J. Hill, P. A. Krasney, B. O'Conner, N. Peirce, and P. L. Greenhaff, "Disuse atrophy and exercise rehabilitation in humans profoundly affects the expression of genes associated with the regulation of skeletal muscle mass," The FASEB journal, vol. 18, no. 9, 2004, pp. 1025–1027.

[6] J. Sweller, "Cognitive load during problem solving: Effects on learning," Cognitive Science, vol. 12, no. 2, 1988, pp. 257–285.

[7] K. R. Koedinger and V. Aleven, "Exploring the assistance dilemma in experiments with cognitive tutors," Educational Psychology Review, vol. 19, 2007, pp. 239–264.

[8] J. Sweller, J. J. Van Merrienboer, and F. G. Paas, "Cognitive architecture and instructional design," Educational psychology review, vol. 10, no. 3, 1998, pp. 251–296.

[9] J. J. Van Merriënboer and J. Sweller, "Cognitive load theory in health professional education: design principles and strategies," Medical education, vol. 44, no. 1, 2010, pp. 85–93.

[10] P. Ayres and T. van Gog, "State of the art research into cognitive load theory," Computers in Human Behavior, vol. 25, no. 2, 2009, pp. 253 – 257.

[11] F. Paas and T. van Gog, "Optimising worked example instruction: Different ways to increase germane cognitive load," Learning and Instruction, vol. 16, no. 2, 2006, pp. 87 – 91.

[12] C. S. Dweck, "Motivational processes affecting learning." American psychologist, vol. 41, no. 10, 1986, pp. 1040–1408.

[13] C. H. Utman, "Performance effects of motivational state: A meta-analysis," Personality and Social Psychology Review, vol. 1, no. 2, 1997, pp. 170–182.

[14] C. Ames, "Classrooms: Goals, structures, and student motivation." Journal of educational psychology, vol. 84, no. 3, 1992, pp. 261–271.

[15] B. E. Vaessen, F. J. Prins, and J. Jeuring, "University students' achievement goals and help-seeking strategies in an intelligent tutoring system," Computers & Education, vol. 72, no. 0, 2014, pp. 196 – 208.

 ISBN: 978-1-61208-390-2

A Dynamic GSOM-based Concept Tree for Capturing Incremental Patterns

Pin Huang, Susan Bedingfield
Faculty of Information Technology
Monash University
Melbourne, Australia
e-mail: phua13@student.monash.edu
e-mail: sue.bedingfield@monash.edu

Damminda Alahakoon
Latrobe Business School
Latrobe University
Melbourne, Australia
e-mail: d.alahakoon@latrobe.edu.au

***Abstract*—The Growing Self Organizing Map (GSOM) has been proposed to address the need of predefining network size and shape in traditional Self Organizing Maps (SOM). In the work described in this paper, the GSOM is used as a foundation for generating hierarchies of concepts in a tree structure which also has the ability to adapt and accumulate new information in an incremental learning architecture. GSOMs are used to capture inputs in time windows and the GSOM nodes are used as the base for developing the bottom level concepts in the tree. A new algorithm is then used to integrate similar information into concepts based on attribute similarities. As new data is introduced, new GSOMs are created and used to capture topological patterns which are integrated into the existing concept tree incrementally. The updated concept tree can capture multiple dimensional inputs with multi-parent nodes. It is proposed that this is an ideal building block to implement the columnar architecture in the human neo-cortex as an artificial model which could then be used as a cognitive architecture for data mining and analysis. The adaptive concept tree model is demonstrated with several benchmark data sets.**

Keywords-growing self organizing map; clustering; concept formation; incremental learning.

I. INTRODUCTION

According to current brain theories, human intelligence and related factors, such as perception, language, prediction, all have a strong relationship to the architecture and structure of the neocortex. The neocortex is believed to be a complex biological auto-associative memory [5], where one of the key features is that patterns from 'experiences' (inputs) are stored in the neocortex in the form of a hierarchy [5]. When storing these patterns, the cortical region provides the group of related active cells a name, and this name is passed to the next higher level in the hierarchy; only the representation of the active cells is passed via the hierarchy; and when the patterns move down the hierarchy, the higher level concepts are broken into granular information [5]. The work described in this paper is based on this base functionality and structure of the neocortex resulting in a model which can capture and accumulate patterns from input data and also adapt to changes with incremental learning. In our proposed concept tree model, lower level represents a more detailed concept and higher level is about a more abstract concept. The information passed from a node at a lower level to a higher level of the tree consist of a median weight value and as such only abstract representative information and no detailed actual information is passed up the hierarchy. This ensures that only high level concepts are captured in the upper levels of the hierarchy.

A further key feature of the neo cortex is that patterns are stored in sequence and activated in sequence with appropriate triggering mechanisms [5]. When we recall our memories, we have to go through it in a sequential order. Although the current version of the proposed model does not demonstrate this functionality, the dynamic and adaptive architecture of the proposed model is an ideal base for developing such capability. This work is currently ongoing as the second phase of the project.

Mountcastle [13] believed that the structure of the neocortex has a columnar organization. The term column can be viewed as a vertical unit in which cells work together. And such columnar unit is the basic computation unit for the cortical computation. The proposed concept tree model is an incremental learning model, which is capable of continuously processing incoming data and adapting as required. The model has the capabilities of generating new columns of sub columns when the new data do not exactly represent past happenings.

The proposed model provides a basis for a larger artificial learning and adaptive model being planned, which can capture accumulate and represent data in a form suitable for decision making. The proposed model is inspired by the current research findings of the neocortex and columnar structure of the brain; therefore, the proposed model embraces some key features: hierarchical concepts formation, incremental learning and adaptation, columnar structure. The GSOM-based tree structure presented in this paper will form an individual column in the larger model with each sub child column representing sub groupings and concepts within each column.

The proposed architecture is made up of three key components: GSOM clustering generated input, a tree base hierarchy, and an incremental update mechanism to accommodate new inputs. Section 2 provides the background for the work described in the paper. The new model and architecture is described in detail in section 3. Experimental results with two benchmark data sets are described in Section 4. Section 5 provides concluding remarks.

 ISBN: 978-1-61208-390-2

II. BACKGROUND

Mountcastle [13] proposed that the structure and appearance of the neocortex is quite uniform and comprises columnar units that run perpendicular to the horizontal layers of the neocortex [13]. The term *column* can be viewed as a vertical unit in which cells work together. Such a columnar unit is the basic computation unit for the cortex's operation. The human neocortex is described as being composed of several hundred millions of mini-columns. Mountcastle [13] also suggested that a cortical area may belong to more than one column or sub column. In other words, a cortical area located in a lower hierarchical level may relate to more than one cortical areas in higher hierarchical levels. This biological feature enables us to relate experiences or inputs to multiple concepts. To accommodate such capability our proposed model enables a child node of a lower level to have more than one parent node of higher levels.

Hawkins [5] has also suggested some key features of the neocortex. For example, patterns are stored in the neocotex in sequence and in the form of hierarchy. Based on his theory of the neocortex, Jeff Hawkins has proposed a Hierarchical Temporal Memory (HTM) model to capture such functionality based on Markov chains [10] and Bayesian belief propagation. These techniques are considered to be symbolic techniques (which deal in human defined abstract symbols) and it has been discussed by Weng [6] that emergent techniques (which can autonomously self-organize via past experience) are more suited to achieving similar functionality to the neocortex. Emergent models include the self-organizing techniques. Our proposed model is based on the GSOM [1].

The GSOM is an unsupervised neural network and has the ability to grow dynamically, the necessity for overcoming the major limitation of the SOM algorithm of a predefined map size. The GSOM algorithm facilitates hierarchical clustering using the Spread Factor (SF) parameter. With a lower SF, a more abstract map can be obtained whereas with a higher SF, a more detailed map can be obtained. In our proposed model, we use a high SF to obtain a very detailed map, which is the building block for the construction of the concept tree. Each node produced from GSOM is viewed as a mini or sub column. In addition, each concept tree which is composed of several hierarchical levels generated from the proposed model, can be viewed as a columnar unit, and its sub trees can be viewed as sub columns. Earlier conceptual clustering models such as CLUSTER/2 [11], do not have incremental learning capability, in contrast, the learning of the human process of incremental knowledge acquisition. There are some incremental conceptual clustering models such as EPAM [3], UNIMEM [9], COBWEB [2], CLASSIT [8], which use different approaches to construct concept trees, however, they do not enable a child concept node to have more than one parent concept node, which means that the model cannot fully implement the neocortex hierarchical structure in which a child node in a column may have more than one parent node located in more than one column.

Lastly, incremental learning related to cognition has been described by Chalup [12] as the development of the brain functionality in three phases. Phase one is the incremental learning that occurs as a result of the evolutionary process over generations. Phase two refers to the neurodevelopment of the brain. This is the stage of acquiring essential abilities such as sensory perception and cognition. Phase three is about the adaptation of the neural system subject to the brain's internal state and the interaction with the environment. Therefore, one of the key features of the proposed algorithm is incremental learning.

III. ADAPTIVE CONCEPT TREE MODEL

A. GSOM

The GSOM algorithm has two modes, the training mode and testing mode. Actual growth of the network and smoothing out of weights occur in the training mode. In the testing phase final calibration of the network occurs if known inputs are used, and for unknown inputs the distance from the existing clusters in the network can be measured.
The training mode consists of three phases. Processing in those three phases is as follows [1].

1) Initializing Phase

a) Weight vectors for the starting nodes are initialized to random numbers between 0 and 1. In general, each map starts with four nodes.

*b) Growth Threshold (GT) is calculated for the given data set based on user requirements. To calculate the GT, the SF parameter value, which is defined prior to the clustering, is used. The formula is GT = -D *ln(SF); here D is the dimension of the input.*

2) Growing Phase

a) Input is presented to the network.

b) The weight vector closest to the input vector is selected using a similarity measuring function. The closest node is considered to be the winner node. The weight vector adaptation takes place for the winner node and the neighbourhood nodes. The amount of adaptation is based on the Learning rate (LR) parameter which is decreased exponentially over the iterations.

c) The error value of the winner node is accumulated by the difference between the winner node's weight vector and the weight vector of the input node.

d) If $TE_i > GT$, where TE_i is the total error value of node i and GT is the Growth Threshold, then new nodes are inserted into the map if node i is a boundary node. If node i is a non-boundary node, the error value is distributed to the neighbourhood nodes.

e) If new nodes are added, weight vectors are initialized to match the neighbouring node weights and initialize the learning rate to the starting value.

f) Repeat the above steps until all inputs are presented to the network and the node growth is set to a minimum level.

3) Smoothing Phase

a) Reduce the learning rate and define a small starting neighbourhood.

b) Present input weight vectors then find winners and adapt their weight vectors and the weight vectors of the neighbourhood nodes in a similar way to the growing phase.

The GSOM algorithm facilitates hierarchical clustering using the SF parameter. SF parameter value is used for the GT calculation and when the SF value is low the GT becomes high, making new node insertion more difficult. In contrast, when the SF value is high the GT becomes low, making new node insertion easier. Because of the above relationship the SF parameter value controls the growth of the output map. Using a lower SF value a more abstract map can be obtained whereas using a higher SF value, a more detailed map can be obtained. This functionality can be used for hierarchical clustering of a given dataset by obtaining an abstract map for the first level of the hierarchy and then further explore the map using a higher SF value.

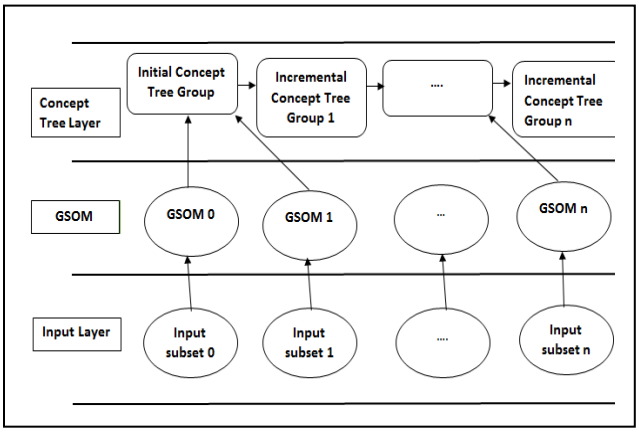


Figure 1. Overall Architecture

B. Overall architecture

The proposed architecture is made of three layers, namely, the input layer, the GSOM layer, and the concept tree layer. This is illustrated in Figure 1. The input layer is where the input data is located. The input dataset can be randomly broken down into several sub datasets. If the input dataset contains temporal features, the input dataset can be broken down by temporality, such that, sub datasets can be organized in a sequential order to represent such temporality. The number of sub datasets should be at least two. When the input dataset has been broken down, the sub datasets will be processed by the model in a sequential order. Furthermore, the number of GSOMs located in the GSOM layer is the same as the number of sub datasets in the input layer. Each GSOM in the GSOM layer is responsible for processing only one sub input dataset. When the first sub input dataset is presented to the first GSOM in the GSOM layer, the output of the GSOM will be presented to the Concept Tree Layer to form the initial concept tree group. After that, the second sub input dataset is presented to the second GSOM in GSOM layer, and then the outcome of the GSOM is presented to the previous established initial concept tree group to form the incremental concept tree group. Similarly, once the sub input dataset has been processed by its corresponding GSOM, the outcome of the GSOM will be presented to previous established concept tree group to generate the next incremental concept tree group.

C. GSOM Layer and Concept Tree Layer Architecture Details

After each GSOM is processed, it presents the clustering results to the bottom level of the previous existing concept tree group, which is illustrated in Figure 2. The concept tree group is composed of three level concept trees (noted as Tree 1 in Figure 2), two level concept trees (noted as Tree2 in Figure 2), and standalone nodes. A standalone node is a level 3 node that does not have any parent nodes at higher levels. Once the bottom level of the concept tree group has processed the input, the information will move up to higher levels. A higher level of the hierarchy means a more abstract concept than a lower level. We set the maximum number of the tree hierarchy to be three; however, the number of hierarchical levels can be set to be more than three by reapplying the same proposed methodology.

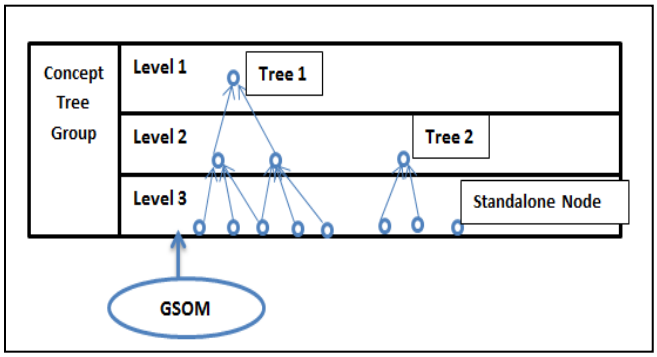


Figure 2. Concept Tree Group

D. Incremental Concept Tree Algorithm

1) Constructing the Initial Concept Tree

Inputs are first presented to the GSOM algorithm. If the value difference of a specific attribute for a pair of nodes agrees to within a predefined value, we say that they have *similar attribute values*. We set this predefined value to be 0.2, which is reasonable because the attribute values are between 0 and 1. Speak of which, attributes' values should be normalized before being presented to the algorithm. In addition, we say that two nodes *share the same concept* if a

predefined percentage of their attribute values are similar. In this paper we use a value of 80% as the predefined percentage. For example, if there are two GSOM nodes (N1 and N2) with m attribute values. N1's attribute values are noted as (A1,A2,…,Am), N2's attributes values are noted as (B1,B2,….,Bm). If the absolute value of (Ai-Bi) is less than 0.2 (here i = 0,1,...,m), we say that N1 and N2 *share similar attribute values* for the ith attribute of the input data. If N1 and N2 share similar attribute values of more than m*20% attributes, we say N1 and N2 share the same concept. Information from the GSOM is first refined then transferred from the GSOM to level 3 (bottom) of the initial concept tree by successively merging the closest node pairs if they share the same concept.

a) Genertating Parent concepts at level 2 for similar nodes at level 3 of the initial concept tree group

For developing concepts from level 3 into level 2 (higher level), two level 3 nodes are defined to be similar if the Euclidean distance between the nodes is less than a predefined distance threshold. We set the threshold as 0.2 * square root of the number of attributes in the input data, which represents the maximum overall distance for all attributes. A parent node of these nodes will be generated at level 2. If a node cannot find any similar node to generate a concept at a higher level, this node will be a standalone node at this level.

b) Generating parent concepts at level 1 for similar nodes at level 2 of the initial concept tree

Similarity between nodes at level 2 is defined in the same way as at level 3. However, because level 1 parent nodes represent more abstract concepts than level 2 nodes, it is appropriate to use a wider distance threshold. We set the distance threshold as 0.4 * square root of the number of attributes of the input data. Parent nodes are created at level 1 for groups of similar nodes at level 2. If a node at level 2 cannot find any similar nodes to form a parent concept at level 1, this node will be without any parent nodes at level 1.

2) Incremental Learning Stage

When the next subset of the input data being presented to its corresponding GSOM, GSOM output nodes are further refined by grouping any closet nodes with similar concepts. Those nodes will be treated as a series of incoming input nodes to level 3 of the already existing concept tree group. If there is no node similar to the input node at level 3 of the existing tree group, the input node will be placed as a standalone node at level 3, which is illustrated in Figure 3.

If the Input node can find similar nodes at level 3, if there is no existing parent node at level 2 able to hold all the child nodes, a new parent node will be created at level 2, which is illustrated in Figure 4. What is more, this enables a child node to have more than one parent node in our proposed model. This new node at level 2 will be treated as an input node to the existing level 2. A similar mechanism of creating a parent node for level 3's nodes is applied to level 2 as well. If the most recently created node at level 2 cannot find any similar nodes at level 2, the node will be added to level 2 as another node. This update will continue up to level 1. Details are illustrated in Figure 5, which is the pseudo code of the incrementally adaptive concept tree algorithm.

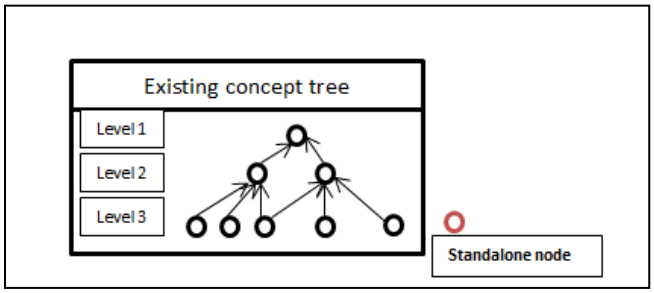


Figure 3. An example of a standalone node at level 3

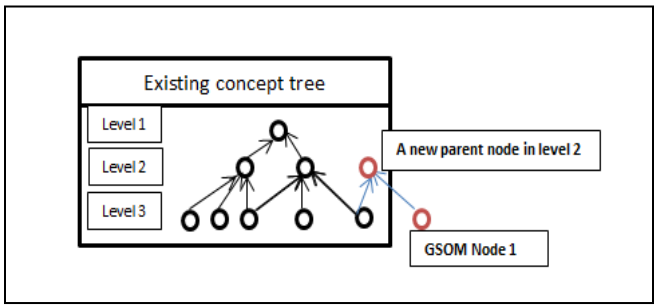


Figure 4. An example of generating a new parent node at level 2

IV. EXPERIMENTAL RESULTS

Experiments were run on two datasets (zoo dataset [7] and heart disease dataset [4]) from UCI data. The zoo dataset is composed of 17 attributes and 101 instances, a majority of attributes are of Boolean type. The Heart disease dataset's attributes are either continuous or Boolean type with 303 instances. The two data sets were chosen to demonstrate the functionality of the new algorithm. The zoo data has been widely used to demonstrate clusters and hierarchical clustering due to the availability of main animal groups and sub groups within. It is also interesting to have animals such as platypus and turtle etc. and see what the algorithm will do with such animals. The key advantage of using this data set is that we can understand why certain animals are grouped together from general knowledge. Also the animal data set has been used to demonstrate the clustering and hierarchical clustering ability of the GSOM and it was the ideal data to show how such clusters are used as a base for concept building and also the incremental update of such concepts. The heart disease data was selected as a more realistic data set, but with attributes which also

 ISBN: 978-1-61208-390-2

has meaning to a general reader. As such it must be emphasized

Algorithm 2 Incrementally Adaptive Concept Tree

When an input node is presented to level 3 of an existing concept tree

Variables:

DistanceThreshold = LevelFactor * square root of the number of attributes of the input data

LevelFactor = 0.2 (level 3)

LevelFactor = 0.4 (level 2)

Procedure Adapting the concept tree in each hierarchical level (Node InputNode)
  Locate a winner node W by computing Euclidean distance.
  If the distance d between InputNode and W <= DistanceThreshold **Then**
    If W does not have a parent node
      Create a node P in a higher level as the parent node of the W and InputNode
      If node P is in level 2 **Then**
        Procedure adapting the concept tree in each hierarchical level (Node P)
      End If
    Else If W has parent nodes
      For Each ParentNode PP in W' Parent nodes
        If InputNode is similar to all PP's Child Nodes **Then**
          Let InputNode be a child node of PP
        ELSE If InputNode is similar to a subset of PP's child nodes **Then**
          Create a new node P in a higher level as the parent Node for them
          If node P is in level 2 **Then**
            Procedure adapting the concept tree in each hierarchical level (Node P)
          End If
        End If
      End For
    End If
  Else
    Add the InputNode to level 3 of the existing tree
  End IF
End Procedure

Figure 5. Incremental Concept Tree Algorithm

that the purpose at this stage is not to evaluate the accuracy of classification of the algorithm, but to demonstrate how GSOM based clusters are used as a base for multi-level concept building with incremental update. At this stage the 'meaningfulness' and 'explain ability' of the concepts are used to evaluate the algorithm. The GSOM has been fully evaluated for cluster accuracy, topology preservation capability and processing advantages. In the following experiments we demonstrate that such GSOM clusters can then be used to develop the concepts which could then be updated as new data changes without losing past learning.

For each node of different hierarchical levels, we calculate the nodes' weighted values and standard deviations for each attribute. These are used to identify the concepts in different hierarchical levels. With the zoo dataset, 16 attributes were used except the last attribute that indicates the animal's category. With the heart disease dataset, null values were removed. Fourteen attributes were used in the experiment, including age, sex and chest pain type, excluding "the diagnosis of the heart disease" attribute. Distinct values of the excluded attribute are 0,1,2,3 and 4, which indicate the probability of having heart disease. The value 0 means absence of heart disease (with less than 50% diameter narrowing), and the value 1,2,3 and 4 stands for different degrees of presence of heart disease (with more than 50% of diameter narrowing). We used a SF of 0.9 to run the GSOM for any subsets of dataset to obtain more detailed maps.

A. Zoo dataset Results

The dataset was divided into two subsets and input to two GSOMs separately. Five concept trees with three levels, six concept trees with two levels, and six standalone nodes were generated from the algorithm.

1) Three level hierachical Concept Tree

The input animals for each concept tree are shown in Figure 6. Tree 0 represent birds, tree 1 is a concept tree for mammals, and tree 2 represents fish. Trees 3, 4 and 5, they all represent reptiles and share some grandchildren (toad, slowworm, and newt).

Top level information provides a general idea about the most abstract concepts. The concept of a node is determined by each attribute's standard deviation and weight values. When an attribute's deviation value is 0, we say that all input instances attached to this node share the same concept, and such concept's name is the attribute's name and the actual value of the concept is determined by the weight values of the attribute. If a node has more than one attribute with zero standard deviation, the concept of a node is the collection of all these attribute's concept. For example, Figure 7 shows Tree 0's top node's attributes' weight values and standard deviations at level 1. The highlighted attributes with 0 standard deviations in Figure 7 stand for the concepts. Animals belonging to Tree 0 share the following concept at level 1: they do not have hair, have feathers, can produce eggs, do not have teeth, have backbones, can breathe, do not have fins, are not venomous, have tails, have two legs – as such birds.

Tree Name	Animals
Tree 0	crow, gull, hawk, kiwi, flamingo, duck, lark, chicken, skua, sparrow,swan,wren, dove, parakeet,pheasant,skimmer
Tree 1	boar,cheetah,leapard,lion,elephant, giraffe,gorilla,calf,lynx,antelope,buffalo, deer,goat,aardvark,bear,mole,opossum, squirrel,vole,mink,pony,pussycat,reindeer, mongoose,wallaby,seal,sealion,polecat, puma,racoon,wolf,oryx
Tree 2	catfish,chub,herring,seahorse,carp,haddock, dogfish,bass,sole
Tree 3	tortoise,toad,newt,slowworm,tuatara
Tree 4	newt, slowworm,tuatara,toad,scorpion
Tree 5	scorpion, newt,slowworm,tuatara,piviper

Figure 6. Three level Concept Tree's input animals

 ISBN: 978-1-61208-390-2

Child nodes inherit their parent nodes' concepts. This is shown in Figure 8. Input instances belonging to Node 3 at level 2 not only share concepts with their parent node at level 1, but also share the concept that animals are not domestic. Similarly, node 9 at level 3 inherits its parents' concepts, and input instances attached to this node also share two more concepts: being predator and not catsize (not the same size as a cat). Therefore, crow, gull, hawk and kiwi are predators and they are not the same size as a cat, they also have the concepts from parent nodes at level 1 and 2. Some nodes at level 3 have more than one parent at level 2 such as Node 7 at level 3. Node 7 and 13 at level 3 inherit the same concepts from their parent (Node 7 at level 2), but they differ in the concept of being domestic or not. Node 7 and Node 8 at level 3 have the same concept inherited from their parent node (Node 2 at level 2), but they differ in the concept of being aquatic or not.

Attribute	Standard Deviation	Weight Values	Attribute	Standard Deviation	Weight Values
hair	0.00	0.00	backbone	0.00	1.00
feathers	0.00	1.00	breathes	0.00	1.00
lay eggs	0.00	1.00	venomous	0.00	0.00
produce milk	0.00	0.00	fins	0.00	0.00
airborne	0.24	0.95	tails	0.00	1.00
aquatic	0.46	0.28	domestic	0.39	0.11
predator	0.48	0.18	catsize	0.33	0.20
toothed	0.00	0.00	legs	0.00	0.25

Figure 7. Tree 0's Standard Deviation and Weight Values

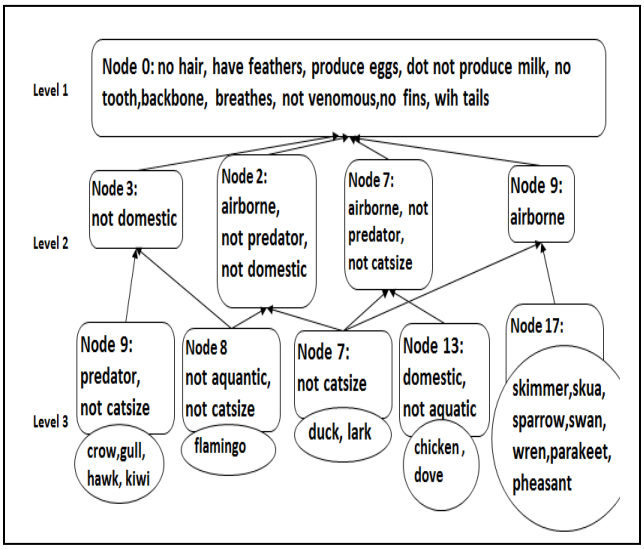


Figure 8. Three Level Concept tree

2) Two level concept trees

These are concept trees which could not be grouped with other nodes to form a more abstract concept at level 1. Figures 9 and 10 illustrate such trees related to aquatic creatures. When compared with the existing three level concept tree, Tree 2 in Figure 6, whose level 1 concept *is no hair, no feathers, produce eggs, no milk, not airborne, aquatic, toothed, backbone, do not breathe, not venomous, fins, tails, no legs.* In Figure 9, octopus, seawasp are not toothed, and some animals in Figure 9 have legs; therefore, this is different from the concept of Tree 2 (no legs and toothed). Similarly, two concepts (milk and catsize) in Figure 10 are different from the concepts in Tree 2; therefore, trees in Figure 9 and 10 cannot be grouped with Tree 2. Figure 11 shows another category different from any concepts in Figure 6. Figure 12's tree shows two animals, cavy and hamster that do not produce milk, however, animals in Tree 1 do produce milk; therefore, they are under different concept trees.

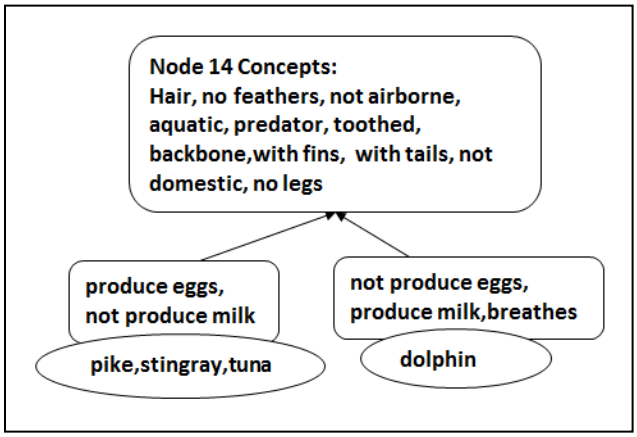


Figure 9. Two Level Concept trees with sea creature 1

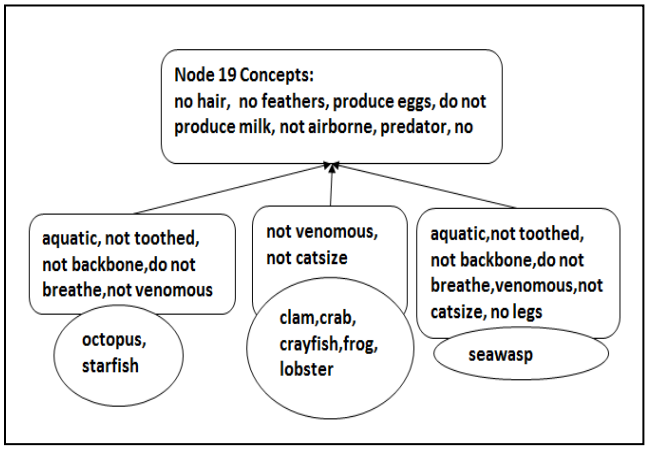


Figure 10. Two Level Concept trees with sea creature 2

3) Standalone nodes at level 3

Figure 13 shows standalone nodes at level.3, which are very different from other animals. Platypus has hair, which is

different from any aquatic animals having parent nodes at level 2 or 3. The seasnake does not produce milk or lay eggs, so it is a sea creature. The fruitbat is an airborne mammal, so it is differs from birds. The ostrich, penguin, rhea and vulture are all big birds. The slug, termite, and worm are not predators, not toothed, and do not have a backbone, therefore, reptiles shown in Figure 6 are different from them.

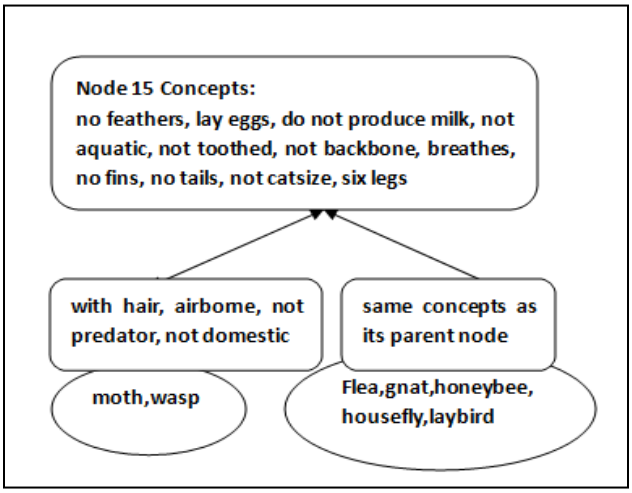


Figure 11. Two Level Concept trees with insects

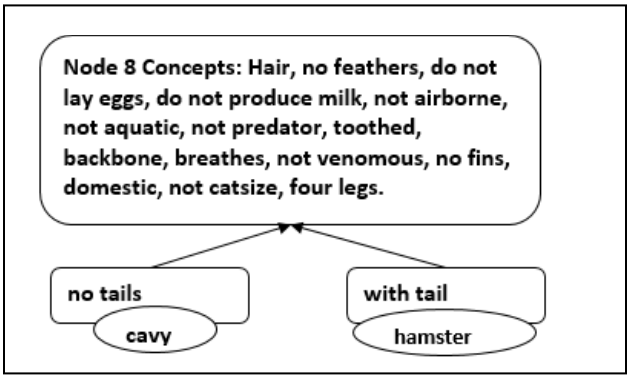


Figure 12. Two Level Concept Tree for cavy and hamster

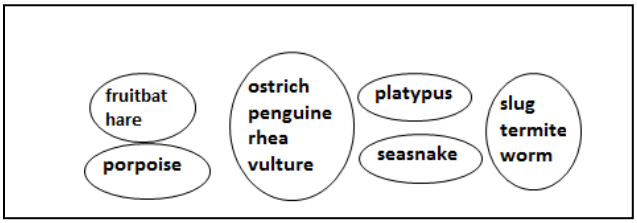


Figure 13. Level 3 standalone nodes

B. Heart Disease dataset Results

1) Three level hierarchical trees

Ten concept trees with three hierarchical levels were created. Figure 14 shows the first level concept in each tree and the percentage of instances belonging to each tree that do not have heart disease. When people do not have anginal pain, more than 80 % of instances under each tree do not have heart disease; when people suffer from anginal pain, it is very likely have heart disease (refer to 18.9% in Tree 0 and 10.4% in Tree 2). Therefore, anginal pain is a very important feature in the diagnosis of heart disease. When a person has anginal pain and "reversable defect", the probability of absence of heart disease increases if they do not have "graphic left hypertrophy". When we analyse concepts from Tree 4 and 5, we can conclude that if people have anginal pain but are "asymptomatic" and "normal (no defect)", their probability of having heart disease decreases compared with instances in Tree 0 and Tree 2. Tree 6, 7 and 8 have 100% of absence of heart disease, showing that when females do not have certain symptoms (indicated in each Tree), they will not have heart disease. From concepts indicated in Tree 6, 7 and 8, we notice that they share some common concepts, such as female, non-anginal pain.

Concept Tree	Tree 0	Tree 1	Tree 2	Tree 3	Tree 4
Level 1 Concept	not atypical angina, anginal pain, no graphic left hypertrophy, reversible defect	non-anginal pain, graphic normal	anginal pain, graphic left hypertrophy, reversible defect	non-anginal pain, graphic left hypertrophy	anginal pain, not asymptomatic , graphic left hypertrophy, Normal (no defect)
Percentage of Absence of Heart Disease	18.9%	81.0%	10.4%	82.1%	72.2%

Concept Tree	Tree 5	Tree 6	Tree 7	Tree 8	Tree 9
Level 1 Concept	female, asymptomatic, not graphic wave abnormality, not graphic left hypertrophy, normal	female, non-anginal pain, zero fasting blood sugar, not exercise induced angina, normal	female, non-anginal pain, not graphic wave abnormality, not graphic left hypertrophy, normal	female, non-anginal pain, not graphic left hypertrophy, not exercise induced angina, normal	asymptomatic, graphic left hypertrophy, normal
Percentage of Absence of Heart Disease	66.7%	100%	100%	100%	35.3%

Figure 14. Content from the three level concept tree

Figure 15 shows that Tree 6, 7 and 8 share some child nodes at level 2. Tree 9's level 1 concept indicates that people with the properties indicated in Figure 14 are more likely not to have heart disease, however, only about 35% of them do not have heart disease. The reason for this is explained by the concept tree as follows. Figure 16 illustrates details of concept tree 9, in which, 'No of Prob_0: 1' means the number of the instances with probability type (the degree of having heart disease) of 0 is one. Similarly, 'No of Prob_1:2' means the number of instances with the

 ISBN: 978-1-61208-390-2

probability type of 1 is 2. Node 45 at level 3 has only one instance. Because of this, we only show concepts that are comparable to sibling nodes' concepts. Node 45 and 64 share the same concepts: zero fasting sugar and zero major vessel, but one group is male, the other group is female. Due to different gender, node 45 and 64 could not be grouped together. Instances in node 45 and 64 are all without heart disease, from which, we can conclude that sex is not significant in determining the presence of heart disease. However, when people do not show any symptoms of chest pain, normal (no defect), zero fasting blood sugar, but have left hypertrophy, they are very likely to not to have heart disease. When we compare nodes 45 and 1, they are all male, but when we compare weight values of the exercise induced angina attribute, instances under node 1 are more likely to have exercise induced angina other than node 45. A majority of people in node 45 have a greater risk of having heart disease, therefore, exercise induced angina is significant in determining the presence of heart disease. This conclusion is further indicated by comparing Node 63 and 1, where instances are all presented with heart disease in Node 64 when they have exercise induced angina, even with zero fasting blood sugar. Another conclusion that can be derived from Node 64 is that fasting blood sugar is not a deterministic feature in determining the presence of heart disease.

2) Two level hierarchical tree

There are six trees with two hierarchical levels. One of the trees is illustrated in Figure 17 where instances have atypical angina, which differs from all level 1 concepts presented in the previous section; therefore, it is reasonable for this tree to not to be grouped with other three hierarchical levels trees. As we can see from the diagram, when people have atypical angina, have zero fasting blood sugar, do not have exercise induced angina and do not have any defects, they are diagnosed with not having heart disease.

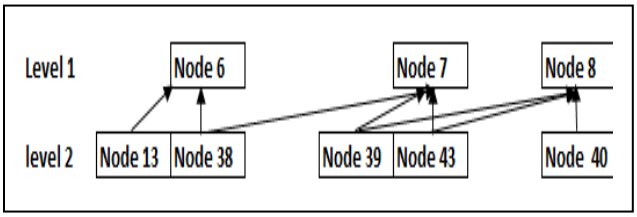


Figure 15. Trees with shared child nodes

3) Standalone nodes at level 3

There are 8 standalone nodes at level 3. For example, Node 4 at level 3, whose concept is "graphic normal", "non-angina pain", "zero fasting sugar", and "reversible defect". 7 out of 8 instances have the value 1 of the attribute "diagnosis of heart disease". When comparing this node with concepts from three level trees, Tree 1's concept (non angina pain, normal graphic and no defect) is quite similar to Node 4. All instances in Tree 1 do not have any defect, which is different from the concept reversible defect in Node 4. That is the reason why Node 4 is not grouped with nodes from Tree 1.

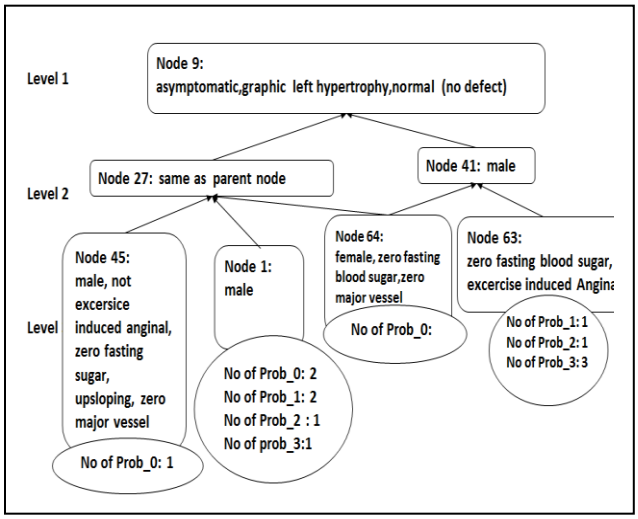


Figure 16. Three Level Concept Tree For Node 9

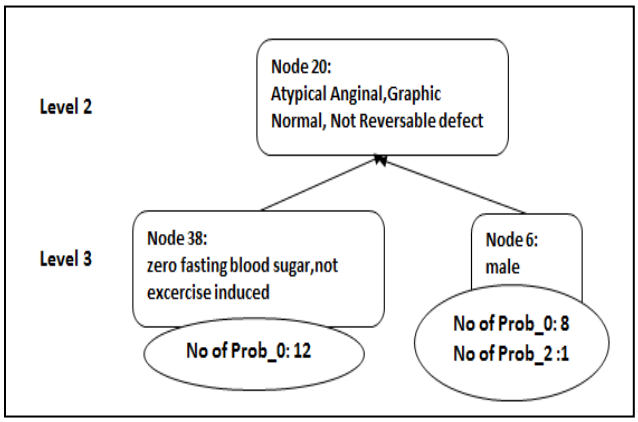


Figure 17. Two level concept tree example

V. CONCLUSION

A new model of information capture, accumulation and adaptation is described in this paper. The model is inspired by the columnar architecture of the neocortex and built using their key concepts and components, Growing SOMs, hierarchical tree structures and incremental learning. The paper describes initial results using two benchmarks data sets from the UCI repository. Although these are not time based data, the input data was divided into subsets and presented in a manner to simulate temporal inputs. The results demonstrate that the model is capable of capturing and representing multi-level concepts from the data and also has the ability to represent sub concepts with multiple parents. This provides the ability of representing a particular situation with multiple 'view points'. The purpose of the presented experiments was not to ascertain the accuracy of classification of the data by the new method. The GSOM

 ISBN: 978-1-61208-390-2

has been utilized with many data sets in the past and has shown to be a useful data clustering and hierarchical cluster generation technique. In the described experiments we use intuitive analysis of the concepts formed by the proposed technique but also have validated these outcomes using past applications of these data sets. But the main focus was the concept formation and incremental update within an architecture based on the columnar formation of the human brain. Such an architecture was essential for the next stage of our research. The described architecture is now being used as the base for implementing cross columnar links and prediction generation. In the current proposed model (which is a key component of the data accumulation and integration model being planned), all the data attributes are processed by the GSOM in GSOM layer, while in the larger complete model, each GSOM component will process a group of relevant attributes, which is the subset of the whole attribute set of the input data. Each GSOM component at GSOM layer will be located in one column. Cross columnar links will be generated to link all columns to demonstrate the inner relationships between columns, which is the foundation for the implementation of the prediction functionality in future complete model. The work is ongoing and the base model described in the paper has provided a good foundation for a dynamic cognitive architecture which could capture sequences in data and also cross columnar relationships in data.

REFERENCES

[1] D. Alahakoon, S. Halgamuge, and B. Srinivasan, "Dynamic self-organizing maps with controlled growth for knowledge discovery", Neural Networks, IEEE Transactions on, vol. 11, 2000, pp. 601-614

[2] D. H. Fisher, "Knowledge Acquisition Via Incremental Conceptual Clustering", Machine Learning, vol. 2, 1987, pp. 139-172

[3] E. A. Feigenbaum and H. A. Simon, "EPAM-like models of recognition and learning", Cognitive Science, vol. 8, 1984, pp. 305-336

[4] D. W. Aha, UCI Machine Learning Repository. Irvine, CA: University of California, School of Information and Computer Science. Available from: http://archive.ics.uci.edu/ml 2014.03.20

[5] J. Hawkins, On Intelligence, 2nd ed, Ameiran: Times Books, 2005, pp. 20-272

[6] J. Weng, "Symbolic Models and Emergent Models: A Review", IEEE Tracnscations on autonomous Mental Development,4, pp. 29-54

[7] K. Bache and M. Lichman, UCI Machine Learning Repository, Irvine, CA: University of California, School of Information and Computer Science, 2013. Available from: http://archive.ics.uci.edu/ml 2014.03.20

[8] M. Gennari, G. Alacqua, F. Ferri, and M. Serio, " Characterization by conventional methods and genetic transformation of Neisserianceae isolated from fresh and spoiled sardines," Food Microbiol, vol. 6, pp. 61-75

[9] M. Lebowitz, "Experiments with Incre- mental Concept Formation: UNIMEM", In Machine Learning, vol. 2, no. 2,1987, pp. 103-138.

[10] J. R. Norris, "Markov Chains", Cambridge University Press, 1999.

[11] R.S.Michalski, " A Theory and methodology of inductive learning," Artificial Intelligence , vol. 20, no. 2, 1983, pp. 111-161.

[12] S. K. Chalup, "Incremental Learning in Biological and Machine Learning Systems", International Journal of Neural Systems, vol. 12, no. 6, 2002, pp. 447-465.

[13] V. B. Mountcastle, " An Organizing Principle for Cerebral Function: The Unit Model and the Distributed System," The Mindful Brain, MIT Press, 1978

 ISBN: 978-1-61208-390-2

Preprocessing of Electroencephalograms by Independent Component Analysis for Spatiotemporal Localization of Brain Activity

Takahiro Yamanoi, Shin-ich Ohnishi,
Faculty of Engineering,
Hokkai Gakuen University,
Sapporo, Japan
e-mail: {yamanoi, ohnishi}@hgu.jp,

Yoshinori Tanaka, Hisashi Toyoshima,
Japan Technical Software Co. Ltd.,
Sapporo, Japan
e-mail: {y-tanaka, toyoshima}@jtsnet.co.jp,

Toshimasa Yamazaki
Faculty of Computer Science and Systems Technology,
Kyushu Institute of Technology,
Iizuka, Japan
t-ymz@bio.kyutech.ac.jp

***Abstract*— The authors measured electroencephalograms (EEGs) from subjects when recalling several types of images. Each image presented consisted of four types of line drawings of body parts. During these experiments, the electrodes were fixed on the scalp of the subjects. However, recorded EEGs had multiple components, including muscle and brain potentials. Recently, independent component analysis (ICA) has been used for EEG analysis. ICA is a technical method for solving the so-called "cocktail party problem". We applied ICA to single-trial EEGs for preprocessing to obtain actual brain activity, and then attempted to estimate spatiotemporal brain activity using equivalent current dipole source localization (ECDL). Our results were almost identical with previous results using ECDL analysis on event Related Potentials (ERPs). In this paper, we present experiments suggesting that ICA is effective as a preprocessing method for estimation of spatiotemporal brain activity using the ECDL and three-dipole model.**

Keywords-Independent Component Analysis; Electroencephalogram; Equivalent Current Dipole Localization Method; Brain Activity.

I. INTRODUCTION

According to research on the human brain, the primary processing of a visual stimulus occurs in V1 and V2 in the occipital lobe. Initially, a stimulus presented to the right visual field is processed in the left hemisphere and a stimulus presented to the left visual field is processed in the right hemisphere. Next, processing moves on to the parietal associative areas [1].

Higher order processing in the brain is associated with laterality. For example, Wernicke's area and the Broca's area are located in the left hemisphere in 99% of right-handed people and 70% of left-handed people [2, 3]. Language is also processed in the angular gyrus (AnG), the fusiform gyrus (FuG), the inferior frontal gyrus (IFG) and the prefrontal area (PFA) [4].

Using equivalent current dipole localization (ECDL) techniques [5] applied to summed and averaged electroencephalograms (EEGs), we previously reported that ECDs can be localized to the right middle temporal gyrus, and estimated in areas related to working memory for spatial perception, such as the right inferior or the right middle frontal gyrus for input stimulus of arrow symbols. Further, using Chinese characters as stimulus, ECD were also localized to the prefrontal area and the precentral gyrus [6 - 10].

To improve the signal-to-noise (S/N) ratio, generally single-trial EEG data are summed and averaged. This results in event related potentials (ERPs). When using such a method, the experiment should be repeated many times to gain accurate data. However, many repetitions can cause subject fatigue.

Recently, independent component analysis (ICA) [11] has been used to analyze acoustic and EEG data. ICA is a mathematical concept used to solve a problem, for example where a number of people are talking simultaneously in a room, and a listener is trying to follow a discussion (cocktail party problem). The human brain can solve this auditory source separation problem, but it is a complex problem for digital signal processing.

Moreover, EEGs used in brain research contain external sources of electric magnetic fields and/or muscle potentials from eye blinks of the subject for example.

In this paper, we used ICA to process single-trial EEGs, and attempted to remove artifacts from the subject's head movements. If effective, then the number of EEG measurement trials could be reduced significantly, that would reduce fatigue of the subject. Furthermore, as we have been using EEGs in a brain-computer interface (BCI) experiment in our laboratory, the use of ICA processing may increase discrimination accuracy.

To confirm the efficiency of ICA, we used an ECDL method for single-trial EEGs after application of ICA [11]. We compared these results with our previous results using ECDL on ERPs [12].

In section II, the ICA methodology is detailed. In section III, preprocessing of EEGs using ICA is described. In section IV, results of one dipole analysis using the ECDL method after ICA are shown. In section V, results of three-dipole analysis using the ECDL method after ICA are shown. In Section VI the results are discussed.

 ISBN: 978-1-61208-390-2

II. INDEPENDENT COMPONENT ANALYSIS

First, we consider unknown n signals generated from independent signal sources

$$s_1(t), s_2(t), \ \dots, s_n(t), \quad (1)$$

and n measured signals that are linearly mixed with the original sources

$$x_1(t), x_2(t), \ \dots, x_n(t). \quad (2)$$

Supposing that a linear relationship with a matrix A, which is independent of time, is

$$x(t) = As(t). \quad (3)$$

It is assumed that the original sources are connected linearly to the signals by a matrix A, as in (3). The purpose of ICA is to separate the signals into n independent components without knowledge of the matrix A, under the assumption of independence of $s(t)$. If a real matrix W exists, then by the following equation

$$y(t) = Wx(t), \quad (4)$$

the original signal can be recovered by reconstructing mutually independent $y(t)$. If $WA = I$ holds then, $y(t)$ coincides with $s(t)$, where I is a unit matrix. However, the equation $WA = PD$ is acceptable, because the order of the signals does not affect the isolation of the components of $y(t)$, where P is a matrix with one single row to each column and D is a diagonal matrix. Under one of the definitions of independence of signals, we determine W so that cost is minimized for one of the cost functions defined elsewhere.

This is the so-called "the cocktail party problem", a mathematical treatment of the effect of discerning the voice of interest from background noise.

III. PREPROCESSING OF EEGS BY ICA

In this study, we used the ICA analysis tool ICALAB (RIKEN Brain Science Institute, Japan) for single-trial EEG preprocessing (Figure 1), which is implemented in MATLAB (MathWorks, Natick, MA, USA). The algorithm we used was Fixed-Point ICA, where correspondence of the input and output is held. For EEG data from 19ch, based on the international 10-20 system, ICALAB is able to obtain the independent components and reconstructs the measured EEG data. In this study, we extracted 19ch component data on ICALAB, an then obtained 19ch EEG data ($y1$, $y2$, …, $y19$) on the basis of the independent components.

Furthermore, as the Fixed-Point ICA generates results for $y1$–$y19$ corresponding to the original 19ch EEG data, we applied the ECDL method to results from ICA to obtained ECDs, and then compared these using a three-dipole model and one-dipole model.

IV. ONE-DIPOLE MODEL OF ECDL AFTER PREPROCESSING WITH ICA

With regard to the EEG data in the present paper, we compared estimated results of the present study with that of our previous results using the ECDL method [12]. Figure 2 shows an example of ERPs obtained from our previous research. The estimated results from multiple brain areas according to latency are shown in Table I. An example of single-trial EEG data is shown in Figure 3, while an example of a single-trial EEG signal processed by ICA is shown in Figure 4.

After preprocessing by ICA, we analyzed the data using the one-dipole ECDL method, and compared the results with Table 1. According to ECD, coincident activity was observed in V1, the posterior central gyrus (PstCG) and the parahippocampus. However, the ICA results revealed no coincident activity estimated in language areas; i.e., Broca's area and Wernicke's area, and the fusiform gyrus, which is involved in cognition. In addition, the goodness of fit (GOF), an evaluation criterion of the estimated results used in the ECDL method, did not reach 99%.

TABLE 1 RELATIONSHIP BETWEEN ESTIMATED ECD AND LATENCY IN EVENT RELATED POTENTIALS FROM MULTIPLE BRAIN AREAS

ESTIMATED PART	V1	ITG	R ParaHip
LATENCY [MS]	131	323	393
R AnG	Wernicke	R Broca	R ParaHip
427	455	485	506
R PstCG	L FuG	R ParaHip	Broca
524	556	590	681

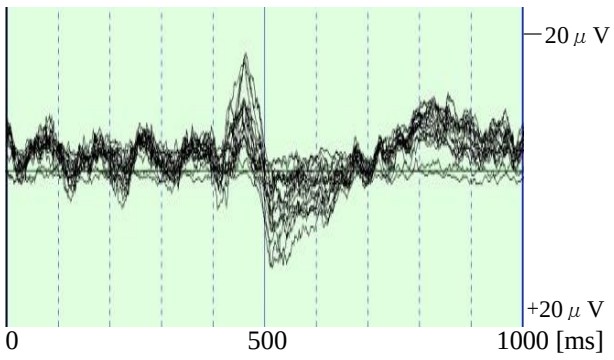


Figure 1. Event Related Potential

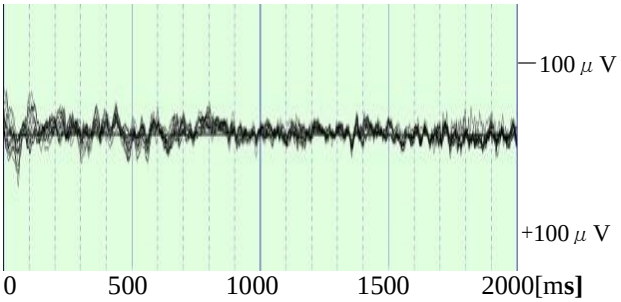


Figure 2. **Example of single trial EEG data**

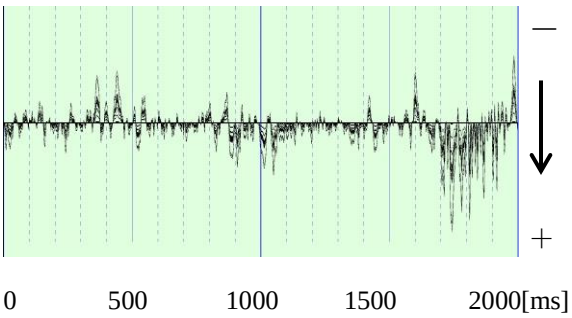


Figure 3. Example of single-trial EEG data processed by ICA (Component *y*1)

In a preanalysis using the one-dipole ECDL method, we found partial concurrence with our previous results [8]. Figure 4 shows an example of estimated ECDs using the one-dipole ECDL method. Two cases are shown, one with disparsed ECDs and the other with concentrated ECDs.

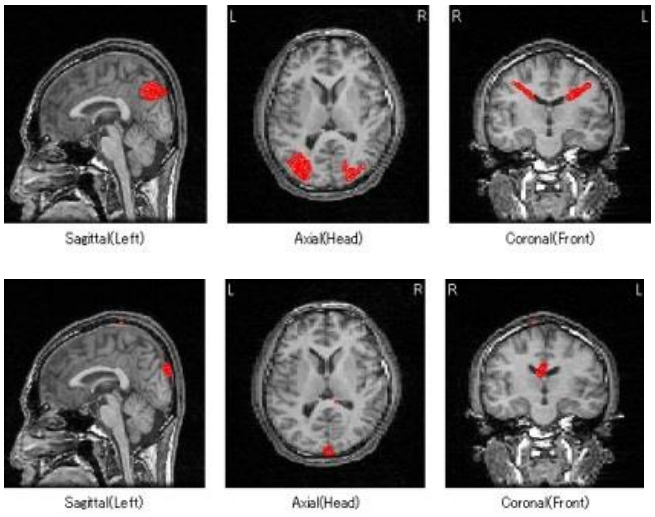


Figure 4. Two Cases of estimated ECDs using the one-dipole ECDL method: Disparsed (Upper) and Concentrated (Lower) ECDs.

To improve estimation accuracy, we applied ICA to a partial data, one half (from 1 to 1000 ms) and one forth (from 1 to 500 ms, and from 501 to 1000 ms). By narrowing the range of latency, the estimated brain parts were increased, but the GOFs were not improved.

V. Application of a Three-Dipole Model of ECDL After Preprocessing Using ICA

Next, we attempted to estimate spatiotemporal brain activity using a three-dipole model, which is more accurate than a one-dipole method, to reconstruct data from *y*1–*y*19 using ICA. In addition, considering the recording of ERPs in Figure 2, we modified the analysis range of latency from 2000 to 1000 ms. Figure 3 shows an example of the results applied to *y*1, among reconstructed data using ICA from *y*1–*y*19. In Figure 3, the sign and ratio is significant in the *y* axis. In addition, Table II shows estimated results of brain areas according to latency. In Table III, we compared the estimated result of each channel *y*1–*y*19 with previous results shown in Table I [8]. Coincident areas are shown in Table III.

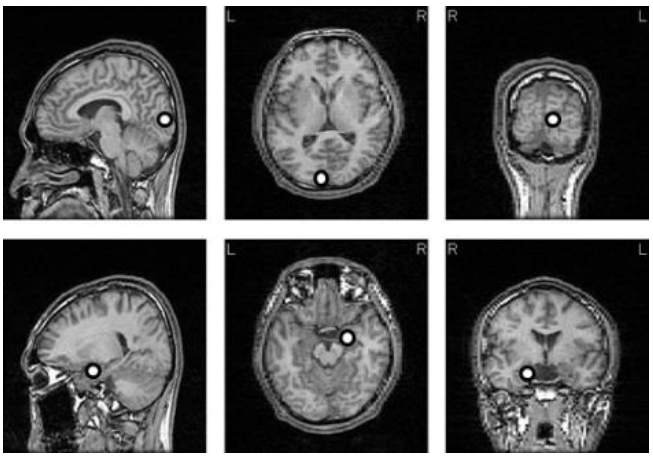


Figure 5. Example of estimated ECDs using the three-dipole ECDL method.

Table II Relationship between Estimated ECD and Latency Using the First Component

Estimated Part	V1	R ITG	R ParaHip
Latency [ms]	128	None	374
R AnG	Wernicke	R Broca	R ParaHip
None	None	None	510
R PstCG	L FuG	R ParaHip	Broca
None	None	595	None

TABLE III ESTIMATED AREAS AND LATENCIES USING THE ECDL METHOD FOR THE 19CH COMPONENT DATA

ESTIMATED PART	V1	R ITG	R ParaHip
LATENCY [MS]	y1: 128 y2: 129 y7: 141 y8: 124	y3: 325 y12: 318 y17: 326 y18: 310	y1: 374 y2: 394 y6: 393 y7: 384 y8: 389 y13: 404 y15: 402 y16: 386
R AnG	Wernicke	R Broca	R ParaHip
y3: 427 y13: 430	y12: 451 y16: 448 y19: 439	y9: 484 y10: 484 y18: 488	y1: 510
R PstCG	L FuG	R ParaHip	Broca
y7: 505 y9: 530	y8: 558 y10: 553 y17: 575	y1: 595 y12: 593 y16: 603	y3: 630 y8: 686 y16: 667

Some coincident results using the one-dipole ECDL method were found and are also shown in TABLE III.

VI. DISCUSSION

In the present study, we compared results from an ECDL three-dipole method to single-trial EEG data, after applying ICA, with results using a similar ECDL method to ERPs.

As shown in TABLE III, our estimated results were almost identical with that of our previous research. However, the new estimated results different in that there were brain areas that were lost in our previous ERP data analyses.

We conclude that ICA is effective as a pre-processing method for estimation using ECDL with a three-dipole model. Moreover, changing the latency range from 0–2000 ms to 0–1000 ms improved the results. These findings suggest that dividing the latency range more precisely may obtain better preprocessing results using ICA for the ECDL three-dipole method used here.

ACKNOWLEDGMENTS

This research was partially supported by a grant from the Ministry of Education, Culture, Sports, Science and Technology for the national project of the High-tech Research Center of Hokkai-Gakuen University March 2013. The experiment was approved by the ethical review board of Hokkaido University.

REFERENCES

[1] R. A. McCarthy and E. K. Warrington: Cognitive neuropsychology: a clinical introduction, Academic Press, San Diego, 1990.

[2] N. Geschwind, and A. M. Galaburda, Cerebral Lateralization, The Genetical Theory of Natural Selection. Clarendon Press, Oxford, 1987.

[3] K. Parmer P. C. Hansen, M. L., Kringelbach, I. Holliday, G. Barnes, A. Hillebrand, K. H. Singh, and P. L. Cornelissen: Visual word recognition: the first half second, NeuroImage, Vol. 22-4, pp. 1819-1825, 2004.

[4] M. Iwata, M. Kawamura, M. Otsuki et al.: Mechanisms of writing, Neurogrammatology (in Japanese), IGAKU-SHOIN Ltd, pp. 179-220, 2007.

[5] T. Yamazaki, K. Kamijo, T. Kiyuna, Y. Takaki, Y. Kuroiwa, A. Ochi, and H. Otsubo: "PC-based multiple equivalent current dipole source localization system and its applications", Res. Adv. in Biomedical Eng., 2, pp. 97-109, 2001.

[6] T. Yamanoi, T. Yamazaki, J.-L. Vercher, E. Sanchez, and M. Sugeno, "Dominance of recognition of words presented on right or left eye -Comparison of Kanji and Hiragana-," Modern Information Processing From Theory to Applications, B. Bouchon-Meunier, G. Coletti and R.R. Yager (Eds.), Elsevier Science B.V., pp. 407-416, 2006.

[7] H. Toyoshima, T. Yamanoi, T. Yamazaki, S. Ohnishi, and M. Sugeno, "Comparison of spatiotemporal activities on human brain for words an symbols with directional meaning (in Japanese)," Journal of Japan Society for Fuzzy Theory and Intelligent Informatics, Vol.18, No.3, pp. 425-433, 2006.

[8] H. Toyoshima, T. Yamanoi, T. Yamazaki, and S. Ohnishi, "Spatiotemporal Brain Activity during Hiragana Word Recognition Task, " J. Advanced Computational Intelligence and Intelligent Informatics, Vol.15, No.3, pp.357-361, 2011.

[9] T. Yamanoi, H. Toyoshima, T. Yamazaki, S. Ohnishi, M. Sugeno and E. Sanchez, "Micro Robot Control by Use of Electroencephalograms from Right Frontal Area," J. of Advanced Computational Intelligence and Intelligent Information, Vol. 13, No. 2, pp. 68-75, 2009.

[10] Y. Tanaka, T. Yamanoi, M. Otsuki, H. Toyoshima, S. Ohnishi, and T. Yamazaki, "Spatiotemporal localization of Brain Activities on Recalling Body Names", 28th Fuzzy System Symposium, 2012, pp.932-935.

[11] A. Hyvärinen, J. Karhunen and E. Oja, "Independent Component Analysis", John Wieley & Sons, 2001.

[12] Y. Tanaka, T. Yamanoi, T. Yamazaki, and S. Ohnishi, "A preprocessing of independant component analysis to spatiotemporal localization of brain activites", Associative Meeting of Societies for Electric and Informatics in Hokkaido Branch, Proceeding in CD, 139, 2013.

 ISBN: 978-1-61208-390-2

Neural Associative Memories as Accelerators for Binary Vector Search

Chendi Yu,
Vincent Gripon
and Xiaoran Jiang

Email: name.surname@telecom-bretagne.eu
Telecom Bretagne, Electronics department
UMR CNRS Lab-STICC
Brest, France

Hervé Jégou

Email: name.surname@inria.fr
INRIA
IRISA, team Texmex
Rennes, France

***Abstract*—Associative memories aim at matching an input noisy vector with a stored one. The matched vector satisfies a minimum distance criterion with respect to the inner metric of the device. This problem of finding nearest neighbors in terms of Euclidean or Hamming distances is a very common operation in machine learning and pattern recognition. However, the inner metrics of associative memories are often misfitted to handle practical scenarios. In this paper, we adapt Willshaw networks in order to use them for accelerating nearest neighbor search with limited impact on accuracy. We provide a theoretical analysis of our method for binary sparse vectors. We also test our method using the MNIST handwritten digits database. Both our analysis for synthetic data and experiments with real-data evidence a significant gain in complexity with negligible loss in performance compared to exhaustive search.**

Keywords–Associative Memories; Binary Sparse Vector; Nearest Neighbors Search; Willshaw Networks.

I. INTRODUCTION

Associative memories are devices that store associations between multiple patterns. They are considered a good model for human memory for their ability to recall stored messages given part of them. For example consider the query of retrieving the word "neuron" from the partially erased query "n*uro*".

The literature on this subject is vast and many models have been proposed, amongst which Hopfield Neural Networks (HNNs) [1] play a prominent role. HNNs store the empirical covariance matrix associated with a set of d-dimensional binary ({-1,1}) vectors. This simple design is appealing. However, HNNs have a strong limitation on the number of vectors they can store. This quantity, referred to as *diversity*, is provably upper-bounded by $d/(2\log(d))$ [2]. Other neural-based methods [3] [4] focus on storing the co-occurrence matrix instead, under the assumption that stored vectors are c-sparse binary ({0,1}). With proper parameters, the diversity of such networks is in the order of the square of d [5].

In machine learning and pattern recognition, many applications aim at matching an input vector to a collection of other ones. In metric spaces, this operation is termed "nearest neighbor search". For high-dimensional vectors, nearest neighbor search complexity is linear in both the number of vectors in the collection and their dimension, as the naive strategy of computing all the distances is the best [6]. To overcome this issue and scale to larger collections, one has to resort to approximate neighbor search techniques, which trade accuracy against scalability [7] [8] [9].

This paper shows that binary neural networks can accelerate nearest neighbor search with limited impact on accuracy, in the case of sparse binary vectors. We support our claim by providing both (i) a theoretical analysis with simulation on synthetic data and (ii) experiments on real data carried out on the gold-standard MNIST handwritten digits database.

The rest of the paper is organized as follows. Section II introduces our method. In section III, we derive a theoretical analysis of the performance, while Section IV presents our real-data experiments. Section V concludes this paper.

II. METHODOLOGY

Consider a set $\mathscr{X}$ of n binary sparse nonzero $\{0,1\}$-vectors with dimension d. Given some input vector x_0, we want to retrieve the closest vector to the query with respect to the Hamming distance:

$$x \in \arg\min_{x' \in \mathscr{X}} d_H(x_0, x') \,,$$

where $d_H(x_0, x')$ denotes the Hamming distance (number of distinct values) between x_0 and x'. This problem is referred to as *nearest neighbor search*.

Since computing distances between two d-dimensional vectors is in $\mathscr{O}(d)$, the complexity of this problem is in $\mathscr{O}(dn)$ with the naive approach, which turns out to be the best choice for exact search: this complexity is tight in high-dimensional spaces. Yet, approximate solutions exist to reduce it [7] [8] [9]. In this case, beyond CPU and memory complexity, the accuracy of the algorithm is a key characteristic when comparing different approaches.

Since in our case, we focus on finding similar neighbors for the Hamming distance and not the Euclidean one, according to Andoni, "the best algorithm for the Hamming space remains the one described in [10]", that is a binary variant of Locality Sensitive Hashing, which requires a lot of memory to be effective in high-dimensional spaces.

The approach we propose below relies on a rather different mechanism (and less memory-demanding) than the hashing-based approach of LSH: neuro-inspired associative memories. It consists of two steps: *selection* and *check*.

 ISBN: 978-1-61208-390-2

a) Selection: We partition $\mathscr{X}$ into subsets $\mathscr{X}_1, \ldots, \mathscr{X}_q$, as

- $\mathscr{X} = \bigcup_{j=1}^{q} \mathscr{X}_j$,
- $\forall j, j', j \neq j', \mathscr{X}_j \cap \mathscr{X}_{j'} = \emptyset$.

In the following, we conveniently consider that the split is regular: $|\mathscr{X}_1| = \cdots = |\mathscr{X}_q| = k$, such that $n = kq$.

We then represent each subset as an associative memory, and search for the input query x_0 in each of them. Formally, we use Willshaw networks. The Willshaw network associated with $\mathscr{X}_j$ is defined as:

$$W(\mathscr{X}_j) = \max_{x \in \mathscr{X}_j} x x^\top .$$

In order to have an estimation of whether some input vector x_0 resembles the content of $\mathscr{X}_j$ or not, we use the following formula:

$$s(x_0, \mathscr{X}_j) \triangleq \frac{x_0^\top W(\mathscr{X}_j) x_0}{(x_0^\top x_0)^2} ,$$

that we shall refer to as the score of x_0 in $\mathscr{X}_j$.

By construction of $W(\mathscr{X}_j)$, the score of x_0 in $\mathscr{X}_j$ is maximal if $x_0 \in \mathscr{X}_j$ and equals one. Conversely, it may happen that the maximal score is achieved even if $x_0 \notin \mathscr{X}_j$, typically when the matrix $W(\mathscr{X}_j)$ is overfilled.

b) Check: Having selected all the associative memories for which the score x_0 is high, we exhaustively compare the query to all the vectors stored in the corresponding subsets. We therefore find the nearest neighbor in a restricted subset of the whole dataset. As we will see theoretically and experimentally, it is likely to contain the desired neighbors to the query.

The complexity of computing the score of x_0 in $\mathscr{X}_j$ is in $\mathscr{O}(d^2)$, and is independent of k. This is the reason that motivates us for using such devices as accelerators for nearest neighbor search. More formally, if the number of subsets that are selected for fine-grain search is N, the overall complexity for performing nearest neighbor search is:

$$C = \mathscr{O}(qd^2 + Nkd) . \tag{1}$$

III. Analysis for synthetic data

A. Theoretical analysis

We consider n binary column vectors x_i of dimension d and L_0 norm $\|x_i\|_0 = c$, c being a constant. The vectors are generated independently and randomly according to a uniform distribution. They are then regularly split by group of k and stored in q Willshaw networks in the way defined in the previous section.

Given an input vector x_0, which is not necessarily contained in the training set, we use this model to perform an approximate search of its nearest neighbor. For simplifying the demonstration, we assume that it has the same norm as the training vectors, $\|x_0\|_0 = c$. Let us denote by x its nearest neighbor and $W(\mathscr{X}_z)$ the matrix it is stored into. We have

$$W(\mathscr{X}_z) = x^\top x + R(\mathscr{X}_z, x)$$

with $R(\mathscr{X}_z, x) = W(\mathscr{X}_z) - x^\top x$.

The score of x_0 in $\mathscr{X}_z$ is written as:

$$\begin{aligned} s(x_0, \mathscr{X}_z) &= \frac{x_0{}^\top W(\mathscr{X}_z) x_0}{(x_0{}^\top x_0)^2} \\ &= \frac{x_0{}^\top x^\top x x_0 + x_0{}^\top R(\mathscr{X}_z, x) x_0}{c^2} \\ &= \frac{\left(c - \frac{d_H(x,x_0)}{2}\right)^2 + x_0{}^\top R(\mathscr{X}_z, x) x_0}{c^2} \\ &\geq \frac{\left(c - \frac{d_H(x,x_0)}{2}\right)^2}{c^2} . \end{aligned}$$

If $x_0 = x$, we have $d_H(x, x_0) = 0$ and $x_0{}^\top R(\mathscr{X}_z, x) x_0 = 0$, thus we rediscover the maximum score of 1.

Let us now consider any other matrix $W(\mathscr{X}_j), j = 1, ..., q, j \neq z$. The score of x_0 in $W(\mathscr{X}_j)$ is calculated as:

$$s(x_0, \mathscr{X}_j) = \frac{\sum_{(l,m) \in M(x_0)} w_{lm}}{c^2}$$

with $M(x_0) = \{(l,m) | x_0(l) = 1, x_0(m) = 1\}$. Obviously, $|M(x_0)| = c^2$.

Let us make the assumption that the random variables w_{lm} are independent. Although this statement is false as a stored vector adds multiple ones in the matrix, it is a fair approximation [4] [5], considering c to be small compared to d. According to the law of large numbers, this normalized sum approaches the expectation $\mathbb{E}(w_{lm})$ when c is large enough. Since w_{lm} is taken within discrete values in {0, 1}, this is nothing else than the probability that any coefficient w_{lm} is equal to 1. We denote this probability $d(W(\mathscr{X}_j))$, and term it the density of the matrix. It can be expressed as:

$$\begin{aligned} d(W(\mathscr{X}_j)) &= \Pr(w_{lm} = 1) \\ &= \Pr(\exists x' \in \mathscr{X}_j, x'(l) x'(m) = 1) \\ &= 1 - \Pr(\forall x' \in \mathscr{X}_j, x'(l) x'(m) = 0) \\ &= 1 - \left(1 - \Pr(x'(l) = 1 \wedge x'(m) = 1)\right)^k \\ &= 1 - \left(1 - \frac{c^2}{d^2}\right)^k . \end{aligned}$$

For sake of simplicity, the matrix that obtains the greatest score is chosen as the winner of this approximate search: $N = 1$. Then, an exhaustive search will be performed for all the vectors that compose the winner matrix. In order to be sure that this matrix contains the nearest neighbor x, the score $s(x_0, \mathscr{X}_z)$ needs to be greater than $s(x_0, \mathscr{X}_j)$ for all $\mathscr{X}_j$. We have then

$$\left(c - \frac{d_H(x,x_0)}{2}\right)^2 \geq c^2 d(W(\mathscr{X}_j)) .$$

If we suppose that $c \ll d$ (the vectors are sparse) and $d_H(x, x_0) \ll c$ (distance from x_0 to its nearest neighbor is small), we obtain:

$$k \leq -\frac{d^2}{c^2} \ln \frac{d_H(x,x_0)}{c} . \tag{2}$$

Let us now analyze the complexity of this method. As given by Equation (1), the complexity when $N = 1$ is expressed as:

$$C = \mathscr{O}(\frac{n}{k} d^2 + kd) .$$

 ISBN: 978-1-61208-390-2

On the other hand, the complexity of an exhaustive nearest neighbor search is written as:

$$C_{\text{knn}} = \mathcal{O}(nd) .$$

An interesting value of k should satisfy:

$$nd \geq \frac{n}{k}d^2 + kd$$

We obtain:

$$\begin{aligned} k &\geq \frac{n - \sqrt{n^2 - 4nd}}{2} \\ &\approx d, \text{if } d \ll n. \end{aligned} \tag{3}$$

Equations (2) and (3) gives the interval of k such that the proposed model finds in expectation the nearest neighbor without error and with a reduced complexity compared to exhaustive nearest neighbor search. The existence of such a value of k implies:

$$-\frac{d^2}{c^2} \ln \frac{d_H(x,x_0)}{c} \geq d$$

We conclude:

$$d_H(x,x_0) \leq \frac{c}{e^{c^2/d}} .$$

For $c = \sqrt{\alpha d}$ with $\alpha > 0$, we have $d_H(x,x_0) \leq \frac{c}{e^{\alpha}}$. Thus, for sparse binary vectors of norm c that grows with the square root of the vector dimension d, if a given vector is sufficiently near from its nearest neighbor, the Willshaw networks can always accelerate the nearest neighbor search without compromising the performance in terms of error rate.

The complexity C should also be compared to that of the Mount Approximate Nearest Neighbor (the Mount ANN) search [7]:

$$C_{ANN} = \mathcal{O}(d(1 + \frac{6d}{\varepsilon})^d \log(n)),$$

where ε is a constant.

In practical scenarios where the dimension is not smaller than 10, it is reasonable to consider $n \ll d^d$. Adding the fact that $d \leq k \leq n$, we easily show that $C = o(C_{ANN})$. Note that in this case, the complexity of the Mount ANN is even larger than that of exhaustive search.

B. Synthetic simulations

We consider a vector set $\mathcal{X}$ that contains $n = 20000$ randomly generated sparse binary vectors of dimension $d = 400$ and $c = 10$ of them are 1s. The test vector is generated as a modified version of any vector in the set $\mathcal{X}$. To generate a test vector x_0, we randomly pick up a vector $x \in \mathcal{X}$ and permute s 1s and 0s, s being smaller than c. Formally, we have

$$\|x \wedge x_0\|_0 = c - s ,$$

and

$$\|x\|_0 = \|x_0\|_0 = c .$$

If s is small enough, x is likely the nearest neighbor of x_0 with Hamming distance $d_H(x,x_0) = 2s$. We take $s = 4$ for the simulations.

Note that we still have not exploited the fact that vectors are sparse in our estimation of algorithm complexity. As a matter of fact, sparsity of vectors helps reducing complexity: we only need to check the positions where are 1s, the number of which is c, rather than looking at every coefficients in a vector of dimension d. Thus, we can update Equation (1) as follows:

$$C = \mathcal{O}(\frac{n}{k}c^2 + \alpha kc) ,$$

and

$$C_{\text{knn}} = \mathcal{O}(\alpha nc)$$

where α is some constant between 1 and 2. In fact, calculating the Hamming distance between two sparse binary vectors of norm c needs at most $2c$ operations if there is not any common position that has coefficient 1, and at least c operations if these two vectors are perfectly identical. We then define the *relative complexity* R_C as the ratio between C and C_{knn}:

$$R_C = \frac{C}{C_{\text{knn}}} = \frac{c}{\alpha k} + \frac{k}{n} \geq \frac{c}{2k} + \frac{k}{n}$$

Therefore, we obtain $R_{C\min} = \sqrt{\frac{2c}{n}} \approx 3.2\%$ in condition that $k = \sqrt{\frac{nc}{2}} \approx 316$ if $n = 20000$ and $c = 10$. Nevertheless, in terms of the error rate, k needs to be small enough to avoid overfitting.

Simulated results are illustrated in Figure 1 with parameters listed as follows: $d = 400$, $c = 10$, $n = 20000$, $s = 4$, $N = 1$. Three measures are assessed in function of k, the number of vectors that contains in each matrix: the red curve represents the Hamming distance gap from the vector that obtains the highest score in our method to the *de facto* nearest vector obtained by the exhaustive search while the blue one indicates the error rate and the black one signifies the relative complexity R_C. For example, when $k = 200$, we obtain a negligible error rate of 0.5% and a Hamming distance gap of 0.003 in consuming only about 3.5% of the complexity of a classical nearest neighbor search.

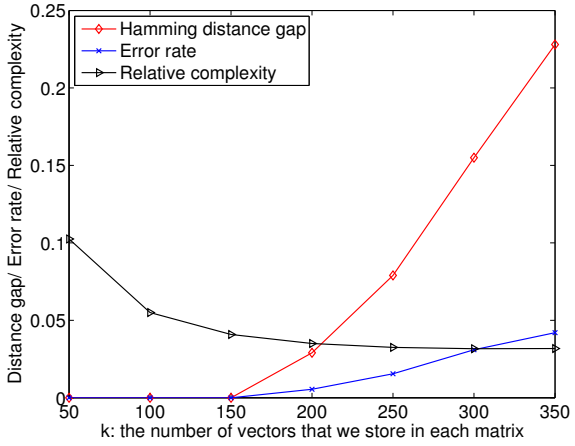


Figure 1. Performance evaluation for the use of associative memory to accelerate nearest neighbor research.

One may notice that, the error rate and the Hamming distance gap grow as k, the number of vectors stored in each matrix increases, which is especially true when k passes

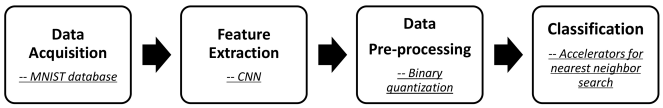


Figure 2. Procedure of the simulation.

a certain threshold (150 in this case). This degradation of performance can be mainly attributed to the saturation of 1s in matrices, which aggravates the noise interference in the selection phase. Since the interference issue is closely related to the matrix density, the performance depends on the length of vectors and the number of vectors contained in each matrix.

IV. Real Data

To evaluate our method on real data, we have carried out experiments on the MNIST handwritten digits database [11]. Created by Yann Lecun, Corinna Cortes and Christopher J.C. Burges, the MNIST database of handwritten digits consists of a training set of 60,000 examples, and a test set of 10,000 examples. All the digits in the database are real-world data.

For the purpose of being more representative, the simulation follows the steps of typical image recognition as is shown in Figure 2.

In the simulation, Convolutional Neural Networks (CNNs) are exploited to extract the proper features followed by a binary quantization. These binary features will be subsequently used as inputs for the proposed method. The nearest neighbor as identified by the proposed method decides which class a test input belongs to. The complexity and the corresponding error rate are then calculated.

For the computation using CNNs, we adapted the code of Rasmus Berg Palm [12]. A synthetic representation of the network considered here is shown in Figure 4 (this representation is inspired by [13]).

More precisely, given some 28×28 image input taken from the MNIST database, we use a 5-layer deep network. The first and the third layers are convolutional layers while the second and the forth layer are sub-sampling layers. The size of kernel for the convolution layers is 5 and the scale of average pooling for sum-sampling layers is 2. Six feature maps are applied for the first two layers and there are 12 for the third and the forth layer. In the fifth layer, we put all the feature maps already present at the fourth layer $(12 \times (4 \times 4))$ into a vector whose dimension is 192×1. The parameter *training rate* is set as 1 and the parameter *training times* as 10.

For the binary quantization, we choose a pre-defined threshold $t = 0.3$. We observe that with this choice the proportion of 1s reaches approximately 0.42 on average in the output vectors. This threshold was tuned to obtain the best performance when using nearest neighbor classification afterwards.

Then we use the proposed method. Let us denote by $\mathscr{X}^D$ the set of transformed training vectors corresponding to digit D, $D = 0,\ldots,9$. For each digit D, we split $\mathscr{X}^D$ into $6000/k$ subsets $\mathscr{X}_j^D$, each of them containing k vectors, and such that $\mathscr{X}^D = \bigcup_j \mathscr{X}_j^D$.

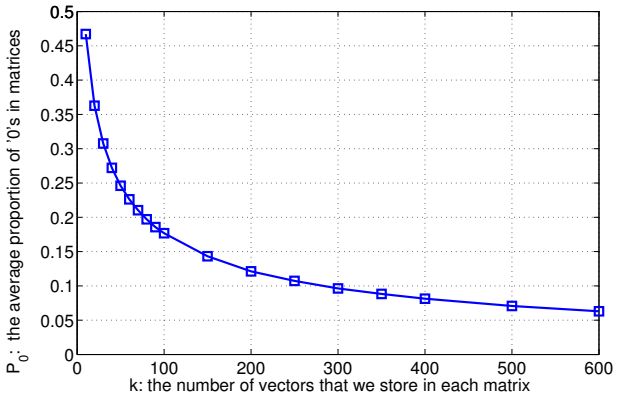


Figure 3. The relation between k and the proportion p_0 of 0s in the matrices.

Therefore, after the training phase, we have $q = \frac{60000}{k}$ matrices in total. Then, at classification time, given a vector x_0 belonging to the test set, we calculate the score of this vector in each set $\mathscr{X}_j^D$, $s(x_0, \mathscr{X}_j^D)$ in the way defined in Section II. The N matrices with the highest scores will be chosen. If there is a tie in matrices scores, we perform a random choice of which to select. The vectors stored in these chosen matrices will be then exhaustively searched to find the nearest neighbor. The label of the matrix that contains the nearest neighbor is returned as the result of the classification process.

The general expression of the overall complexity given by Equation (1) should be adapted in this real data scenario. In fact, the matrix $W(\mathscr{X}_j^D)$ contains much more 1s than 0s. Clearly, as k grows, the number of 0s in the matrix decreases. However, due to the non-uniformity of the vectors in MNIST database, the speed at which 0s decreases is not as fast as for synthetic data (see Section III). Figure 3 depicts the proportion of 0s p_0 in the matrices as function of k.

Therefore, we can exploit this fact to adjust the computation of complexity. In fact,

$$s(x_0, \mathscr{X}_j^D) = \frac{x_0^\top W(\mathscr{X}_j^D) x_0}{\left(x_0^\top x_0\right)^2} = 1 - \frac{\sum_l \sum_m x_{0l} x_{0m}}{\|x_0\|^2}$$

the values of l and m satisfying $w_{lm} = 0$.

In this way, we only have to look at the positions where are 0s in the matrix $W(\mathscr{X}_j^D)$. The overall complexity becomes:

$$C = \mathscr{O}(p_0 q d^2 + N k d)$$

Simulations are performed using a wide range of values of k and N. In Figure 5, the error rate is represented in function of the relative complexity compared to exhaustive nearest neighbor search. We observe that the complexity can be greatly reduced (up to a factor of 0.3 compared to the exhaustive search) for almost the same error rate when parameters are well chosen. The results we obtain in terms of error rate are similar to that obtained using the Mount ANN [7].

 ISBN: 978-1-61208-390-2

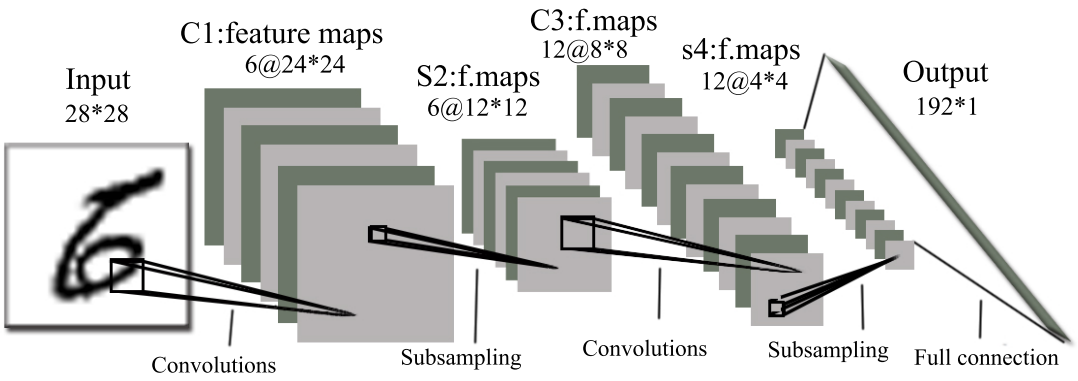


Figure 4. Representation of the CNNs layers used to extract features from MNIST raw images.

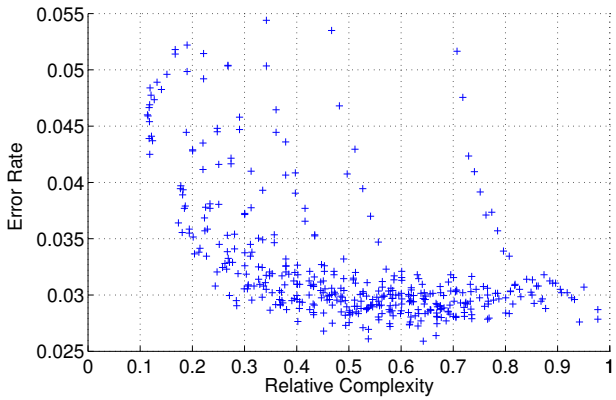


Figure 5. Error rate in function of the relative complexity. Simulations are performed for the MNIST handwritten digits database.

V. Conclusion

We proposed a method to perform approximate nearest neighbor search using neural-based associative memories. This method advantageously takes benefit from the fact that the neural network dimensions of the representation associated with a set of k vectors is independent of k.

As a result, we prove that for synthetic data, one can expect dramatically lower complexity with no reduction on accuracy. We also evaluate our method using real data and prove that it is possible to achieve a reduction of 70% of complexity with a limited impact on the error rate.

In future work, we consider looking at alternative ways to use Willshaw networks to really benefit from their associative capabilities [14], as well as perform experiments on more challenging datasets.

Acknowledgment

This work was founded in part by the European Research Council under Grant ERC-AdG2011 290901 NEUCOD.

References

[1] John J Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proceedings of the national academy of sciences*, vol. 79, no. 8, pp. 2554–2558, 1982.

[2] Robert J McEliece, Edward C Posner, Eugene R Rodemich, and Santosh S Venkatesh, "The capacity of the hopfield associative memory," *Information Theory, IEEE Transactions on*, vol. 33, no. 4, pp. 461–482, 1987.

[3] David J Willshaw, O Peter Buneman, and Hugh Christopher Longuet-Higgins, "Non-holographic associative memory.," *Nature*, vol. 222, pp. 960–962, 1969.

[4] Vincent Gripon and Claude Berrou, "Sparse neural networks with large learning diversity," *IEEE Transactions on Neural Networks*, vol. 22, no. 7, pp. 1087–1096, July 2011.

[5] Judith Heusel, Matthias Löwe, and Franck Vermet, "On the capacity of a new model of associative memory based on neural cliques," *arxiv preprint*.

[6] Roger Weber, Hans-J. Schek, and Stephan Blott, "A quantitative analysis and performance study for similarity-search methods in high-dimensional spaces," in *Proceedings of the International Conference of Very Large Databases*, 1998, pp. 194–205.

[7] Sunil Arya, David M Mount, Nathan S Netanyahu, Ruth Silverman, and Angela Y Wu, "An optimal algorithm for approximate nearest neighbor searching fixed dimensions," *Journal of the ACM (JACM)*, vol. 45, no. 6, pp. 891–923, 1998.

[8] Sunil Arya, David M Mount, Nathan S Netanyahu, Ruth Silverman, and Angela Wu, "An optimal algorithm for approximate nearest neighbor searching," in *Proceedings of the fifth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics, 1994, pp. 573–582.

[9] Piotr Indyk and Rajeev Motwani, "Approximate nearest neighbors: towards removing the curse of dimensionality," in *Proceedings of the thirtieth annual ACM symposium on Theory of computing*. ACM, 1998, pp. 604–613.

[10] Aristides Gionis, Piotr Indyk, and Rajeev Motwani, "Similarity search in high dimensions via hashing," in *Proceedings of the International Conference of Very Large Databases*, 1999, pp. 518–529.

[11] "Mnist database," 2014, URL: http://yann.lecun.com/exdb/mnist/ [accessed: 2014-01-02].

[12] "Deeplearntoolbox," 2014, URL: https://github.com/rasmusbergpalm/DeepLearnToolbox [accessed: 2014-01-02].

[13] "Net5," 2014, URL: http://eblearn.sourceforge.net [accessed: 2014-01-02].

[14] Ala Aboudib, Vincent Gripon, and Xiaoran Jiang, "A study of retrieval algorithms of sparse messages in networks of neural cliques," in *Proceedings of Cognitive 2014*, May 2014, pp. 140–146.

 ISBN: 978-1-61208-390-2

Induction of Intentional Stance in Human-Agent Interaction by Presenting Goal-Oriented Behavior using Multimodal Information

Yoshimasa Ohmoto*, Jun Furutani* and Toyoaki Nishida*
*Department of Intelligence Science
and Technology
Graduate School of Informatics
Kyoto University
Kyoto, Japan
Email: ohmoto@i.kyoto-u.ac.jp, jfurutani@ii.ist.i.kyoto-u.ac.jp, nishida@i.kyoto-u.ac.jp

***Abstract*—We have made noticeable progress in developing robots and virtual agents; human-like robots and agents are closer than ever to becoming a reality. We want to develop an embodied conversational agent that is regarded as a social partner, not just multimodal interface. However, the mental stance of people when they interact with agents is usually different from when they interact with humans. Therefore, in some cases, it is difficult for people to speculate on an agent's emotion and it is also difficult for an agent to persuade people. To solve this problem, we focused on "intentional stance". Intentional stance is a mental state in which we think that an interaction partner has intention. We hypothesized that agents could induce the intentional stance by performing goal-oriented actions in human-agent interaction. To investigate the effect of induction of intentional stance, we made two agents: a "trial-and-error agent" that performed goal-oriented actions using multimodal behavior and a "text display agent" that displayed its behavioral intention via text. We conducted an experiment in which two participants played customized tag in virtual reality with one of the agents. The results showed that participants continuously tried to communicate with the trial-and-error agent, which did not respond to the participant's actions except when necessary for performing the task. We found that the participants felt that the agent using multimodal nonverbal behavior was more goal-oriented, more intelligent and understood their intentions more than the agent that displayed text above its head. Thus, we were able to induce the intentional stance by presenting a trial-and-error process using multimodal behavior.**

Keywords–Multi-modal interaction, human-agent interaction, intentional stance.

I. INTRODUCTION

In recent years, noticeable progress has been made in developing Embodied Conversational Agents (ECAs), such as robots and virtual agents. Human-like ECAs are closer than ever to becoming a reality. We want to develop an ECA that is regarded as a social partner, rather than just a multimodal interface. Many issues must be dealt with in the production of a social partner agent, such as flexible conversation ability, learning ability in novel situations and so on. We focus here on the issues that relate to the construction of human-agent relationships.

Roubroeks et al. [1] reported the occurrence of psychological reactance when artificial social agents are used to persuade people. In that study, participants read advice on how to conserve energy when using a washing machine. The advice was either provided as text-only, as text accompanied by a still picture of a robotic agent, or as text accompanied by a short film clip of the same robotic agent. The results of the experiment indicated that the text-only advice was more accepted than either advice with the still picture of the robotic agent or the advice with the short film clip of the robotic agent. Social agency theory proposes that more social cues lead to more social interaction, but the result was the exact opposite. This is caused by differences in people's mental state with respect to humans or agents. These differences provide a critical barrier for an ECA to cross before it can be accepted as a social partner. It is thus important that the mental state of people when they interact with the agents is the same as that when they interact with humans.

The mental states that humans can be in with respect to an agent can be defined as physical stance, design stance and intentional stance [2]. When we take the physical stance, we pay attention to physical features, such as the power of the motor, the spec of the display and so on. When we take the design stance, we expect that the agent works mechanically according to predefined rules. When we take the intentional stance, we consider that the agent has subjective thoughts and intentions. When a human interacts with another human, they usually take the intentional stance. In this case, they and their communication partner respect each other. When a human interacts with a machine, they usually take the design stance. In this case, they usually interact with the machine from a self-centered perspective because they do not consider that the machine has its own intentions. To establish social relationships between a human and an artificial agent, the agent has to induce the intentional stance.

The purpose of this study is to investigate how to induce the intentional stance in human-agent interaction. The final goal is to establish social partner relationships between humans and agents. For this purpose, we propose a method to induce the intentional stance, implement the method in an agent and experimentally investigate the effect of inducing the intentional stance.

The paper is organized as follows. Section 2 briefly introduces previous work on the intentional stance. Section 3 explains the outline of the proposed method to induce the intentional stance. Section 4 describes an experiment for comparing two types of methods and then presents the results. Section 5 discusses the achievements and limitations. Section 6 concludes and discusses future work.

 ISBN: 978-1-61208-390-2

II. Related work

Heider and Simmel [3] demonstrated that observers attribute elaborate motivations, intentions and goals to even simple geometric shapes based solely on the purposeful pattern of their movements. Dittrich and Lea [4] discussed that the perception of object's motion as animate depended not only on the interaction between the objects, but also on goal-oriented behavior conveyed by it. From these studies, it can be concluded that goal-oriented behavior is important in the induction of the intentional stance.

If an agent resembles a human or an animal in appearance, people tend to spontaneously think that the agent has intentions. Friedman et al. [5] reported that 42% members of discussion forums about a animal robot named AIBO, a robotic pet, spoke of AIBO having intentions or that AIBO engaged in intentional behavior. On the other hand, some people think that AIBO is just programmed robot. They usually get bored with interacting with AIBO in a short time. These show that the mental stance can dynamically change throughout interaction.

In this study, we attempt to induce the intentional stance by presenting goal-oriented behavior. In short interactions, people take the intentional stance when the agent has similar appearance to a human. However, we aimed at long-term interaction because our final goal is to establish social partner relationships between humans and agents. Therefore, we evaluated how successful the induction of the intentional stance was after a certain length of interaction.

Chen et al. [6] reported that the perceived intent of the robot significantly influenced people's responses when a robotic nurse autonomously touched and wiped each participant's forearm. The participants responded less favorably when they believed the robot touched them to comfort them versus when they believed the robot touched them to clean their arms. In this study, they used the robot's speech, the actions of its arm, and the nursing scenario to convey intent of the robot. These explicit cues could quickly induce intentional stance. We expect, however, that the affective relationship between participants and robot may be short-lived because they can easily estimate the mechanisms of the robot behavior and they feel that the robot mechanically interacts with them.

III. A method for presenting goal-oriented behavior

Recognizing the goal-oriented actions of the artificial agent is important for taking the intentional stance. There are many ways to present goal-oriented actions, but, we think, they are not always useful in inducing the intentional stance. For example, optimized actions for a particular goal are goal-oriented actions, but we do not tend to think that an optimized agent has human-like intentions. In this study, we propose two methods for presenting goal-oriented actions: showing a trial-and-error progression towards a goal using multimodal behavior, and displaying the agent's behavioral intention using text. We compared the differences between these methods and investigate how effective they are at inducing the intentional stance.

A. Task description

In this study, we used a "customized tag" game in virtual space as an interaction task. Some rules were added to the rules of normal tag, such as "a tagger cannot tag to players who stand on higher place than the tagger's place," "a tagger can tag to players on higher place after the tagger stops in front of one of the players on higher place and counts to five," "players can only move limited area separated by the virtual water." The virtual game environment did not automatically controlled. This means that the players (two humans and one agent) themselves had to judge and communicate about whether the tagger had changed or not, whether the "five count" was finished or not and whether the players moved the valid region. The game settings encouraged players to consider different objectives to enjoy the game, such as chasing a fastest runner as often as possible, forcing all of the players to be a tagger at least once and so on.

When people play a playground game like a tag game, each player has a different objective in their enjoyment of the game. To ensure that all players enjoy the game, it is important to understand each other's different objectives or goals. Therefore, when playing the customized tag game, participants can take the intentional stance depending on the behavior of the playing partners. In addition, using this game for an experiment allows us to obtain good data for analysis because participants become quite involved in the game [7].

In the customized tag game, the players actively communicate with each other to make all players enjoy the game, such as discussing about the way to control the game (e.g., how to judge the tagger change), advising other players (e.g., "wait! wait!" and "please chase other player!") and seeking to approval (e.g., "he is cheat! I cannot go to the place!"). The communication behavior is not expressed to non-player characters in general video game. So, we focused on the communication behavior to evaluate whether the intentional stance was induced or not.

B. Outline of architecture

The agents in this study decide their behavior through three layers: a goal layer, a behavior category layer and a concrete behavior layer. The elements of each layer are predefined by a designer of the agent.

The goal layer is the most abstract layer. The elements of the layer show the task goal. In our task, this layer has three goals: chasing other players for an extended time, making the time during which a player is a tagger equal among all the players, and making the number of times that a player becomes a tagger equal among all the players. The first goal is predefined, but when other players do not accept the goal, the goal is changed depending on the other players' behavior.

The elements of the behavior category layer show the category of possible behavior. In this study, the categories of behavior include "chasing", "provocation", "dissatisfaction", "escape" and "hiding". Each category has a parameter named "effect level", which indicates how effective it is at achieving a selected goal in the goal layer. Each category is a subgoal of a concrete behavior.

The elements of the concrete behavior layer show the concrete behavior produced by an agent. Each concrete behavior has a parameter named "expression strength", which indicates how clearly the behavior expresses the subgoal of the behavior.

The outline of the system architecture is shown in Figure 1 and was developed based on a Belief-Desire-Intention (BDI) model. The overview of each component is briefly explained below.

 ISBN: 978-1-61208-390-2

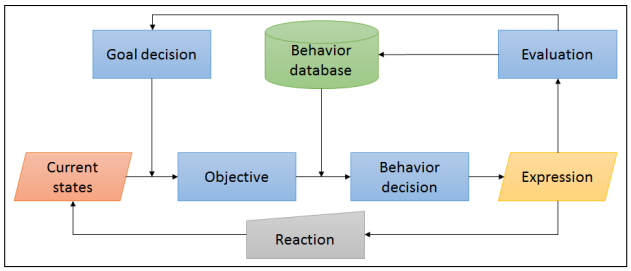


Figure 1. The outline of the system architecture.

Current states:
These are the inputs to the system. The inputs include who is the tagger, each player's position, the time taken to chase other players, how long a player has been a tagger and so on.

Goal decision:
This component determines which goal to achieve. The goal is selected in the goal layer based on the predefined rules.

Objective:
This component determines a category of the behavior. The category is selected in the behavior category layer. This component also determines the values of the "effect level" and the "expression strength" that are needed to achieve the goal.

Behavior database:
The database contains all of the possible behaviors and the structure of the behaviors that are defined in the behavior category layer. The database also contains the current "effect level" and the predefined "expression strength".

Behavior decision:
This component decides a concrete behavior based on the received values of the "effect level" and the "expression strength".

Expression:
This component produces the selected behavior.

Evaluation:
This component evaluates the effect of the concrete behavior on achieving the selected goal. The current values of the "effect level" in the behavior database depend on the evaluation.

C. Method 1: presenting a trial-and-error process using multimodal behavior

In this method, we present a trial-and-error process of achieving a goal using multimodal behavior, such as hand gestures, body orientation, moving speed and iconic motions. Here, we hypothesize that the mental stance, such as the design stance and the intentional stance, changes depending on the agent behavior estimation model. People construct a behavior estimation model but imperfect through interacting with the agent with this method, because it is difficult to precisely interpret multimodal behavior in terms of estimating the behavioral goal. Since we expect consistent goal-directed behavior from the agent, people regard any uncertainties as being caused by the "intentions" of the agent.

The agent uses both the "effect level" and "expression strength" parameters. The expression strength parameter is rated from one to four for each concrete behavior by the designer of the agent behavior, but its value does not change during the task. The value of the effect level changes depending on the output of the evaluation component.

The agent selects its behavior category depending on the value of the effect level. When the same behavior category is selected and achieves the subgoal of the concrete behavior, the agent produces a concrete behavior with higher expression strength than before. When the agent is less able to achieve the selected goal, the value of the effect level decreases. When the effect level of the particular category is less than that of other categories, the behavior category is changed. For example, when the agent wants to provoke a tagged player, first action is "standing near the tagged player." After the provocation is contributed a selected goal (for example, provocation often encourages extending the time to chase other players), the action which has greater value of "effect level" is selected in next phase of the game, such as "waving hand near the tagged player" or "jumping and waving hand near the tagged player." Of course, if the action does not contribute the goal, the agent changes the behavioral category and tries to encourage the goal.

Changes in the expression and the behavior category are thus made in a trial-and-error fashion to achieve the goal. Therefore, when they observe this kind of trial-and-error process using multimodal behavior, people construct a behavior estimation model containing some uncertainties.

D. Method 2: displaying the agent's behavioral intention using text

This method encourages the construction of a behavior estimation model of an agent by displaying intentional agent behavior via text. In this study, the intention of the agent's behavior is a category of the behavior, because each category is a subgoal of a concrete behavior. People construct a behavior estimation model with no black boxes through interacting with the agent with this method, because they can precisely understand the intention of the agent. If only presenting goal-oriented behavior is important in taking the intentional stance, then people interacting with the agent with this method should take the intentional stance.

The agent produces patterns of text corresponding to the behavior category. The diversity of the representations of goal-oriented behavior is the same as the method of presenting a trial-and-error process using multimodal behavior. For example, when the agent wants to provoke a tagged player, the agent displays one of the text expressions, such as he is waiting your chase, he is relaxing and a little bored, "you can't catch me" or "are you tired?" The expression strength is not rated in each text and the text in the same category is randomly selected when the concrete behavior is determined. The value of the effect level changes depending on the output of the evaluation component. The category of the behavior changes in the same way as in the trial-and-error agent.

IV. EXPERIMENT

To investigate the effect of inducing intentional stance, we conducted an experiment using two agents: one was a "trial-and-error agent" that performed goal-oriented actions using multimodal behavior and the other was a "text display agent" that displayed its behavioral intention using text. These agents

 ISBN: 978-1-61208-390-2

were controlled manually (Wizard of Oz) but the behavior planning of the game and the expressions of the multimodal behavior and the text were automatically controlled. We used a virtual reality "customized tag task".

To evaluate the effect, we asked the participants to answer questionnaires after the experiment. In addition, we analyzed the number of communicative actions towards the agent throughout the experiment. Since both agents do not respond to the participant's actions except when they need to respond to allow them to perform the task (for example, when the participant is chasing the agent and when the participant argues that the agent is tagged), the number of communicative actions of participants towards the agent decreases. However, we consider, when the participants have intentional stance towards the agent, they unconsciously try to communicate with the agent in the same ways with humans (e.g., calling the agent's name when they excited in the game, waving hand towards the agent who was chased by other player, asking the way to go to the place near the agent and offering a particular action towards the agent). We focused on such communicative actions. The communicative actions were annotated by two annotators and we adopted communicative actions which was annotated by both annotators. We compared the experimental results between a group in which participants interacted with the "trial-and-error agent" and a group in which participants interacted with the "text display agent."

A. Task

Two humans and an agent (which randomly selected the trial-and-error or text display behavior) participated in the customized tag task. We used the virtual space showed in Figure 2, and added a rule to limit the movement range using a region of virtual water.

The game was not controlled automatically. The players (the humans and the agent) judged whether the chaser had changed, whether the count to five had finished and whether the players moved to a valid region on their own.

The human players were allowed free communication using verbal and nonverbal information. Both agents only communicated when it was necessary to perform the task. They did not respond to the utterances of human players in other situations.

B. Experimental setup

In this study, we used Immersive Collaborative Interaction Environment (we call this ICIE) [8] and Unity[9] to construct the virtual environment and the two agents. ICIE uses a 360-degree immersive display that is composed of eight portrait orientation monitors with a 65-inch screen size in an octagonal shape. In this environment, participants could easily look around in the virtual space with low cognitive load like in the real world. The player's virtual avatar could be controlled by their body motions using a motion capture system embedded in the ICIE. The participants could thus easily interact using body motions with low physical constraints. To move in the virtual space, the players used the Wii controller to move the virtual environment. The controller did not interfere with the player's body motions because it was lightweight and had a wireless connection.

Two video cameras recorded the participants' behavior; one was placed on the screen facing the participants, and another was placed behind them. The participants' voices were recorded using microphones. All of the input and output in the virtual space were also recorded.

C. Participants

Sixteen students (14 males and 2 females) participated in the experiment. They were undergraduate students from 18 to 25 years old (an average of 21.5 years old). All of them interacted with one of the agents for 40 minutes. Eight participants (7 males and 1 females) interacted with the trial-and-error agent and the rest interacted with the text display agent. The experimenter gave the following instructions about the agent: "the agent can recognize your speech. The agent has a lot of knowledge about the customized tag task." We expected that the participants thought that the agent had conversation ability at least at the beginning of the human-agent interaction.

D. Results

To investigate the degree of induction of the intentional stance, we analyzed the number of communicative actions towards the agent and the participants' subjective impressions of the agent using questionnaires.

1) Analysis of the number of communicative actions towards the agent: The purpose of this analysis is to investigate whether performing goal-oriented behavior influenced the actual communication behavior related to the intentional stance. For this purpose, we counted the number of communicative actions towards the agent. We expected that the number of the actions would decrease when a participant took the design stance because he/she think that the agent never react to his/her communicative actions. On the other hand, we would consistently observe actions because the participant unconsciously produced the communicative actions when they took the intentional stance.

To analyze the changes in the number of the communicative actions throughout the experiment, we divided the time series of the experiment evenly into four periods and counted the number of communicative actions in each period. After that, we compared the number in the second period with that in the fourth period. We did not use the number in the first period because during this period the participants were still learning how to control their avatars and how to play the game. In other words, the first period was the "ice breaking" period.

T-tests were used on the data from the trial-and-error agent and the text display agent for comparing the numbers in the second and fourth periods. The results are shown in Figure 3: the number in the fourth period was significantly less than that in the second period only for the text display agent ($p = 0.0003$). The participants could clearly estimate the behavior model of the text display agent and in the end they took the design stance towards this agent. The number in the second period for the text display agent was more than that for the trial-and-error agent (but there was no significant difference). We assume that the participant could easily estimate the goal of the text display agent's behavior because its intention was clear. From these results, we suggest that clearly presenting the goal of the behavior can quickly induce the intentional stance, but that the stance quickly changes to the design stance because humans can construct a precise behavior model.

On the other hand, three participants out of eight increased the communicative actions in the trial-and-error agent group

 ISBN: 978-1-61208-390-2

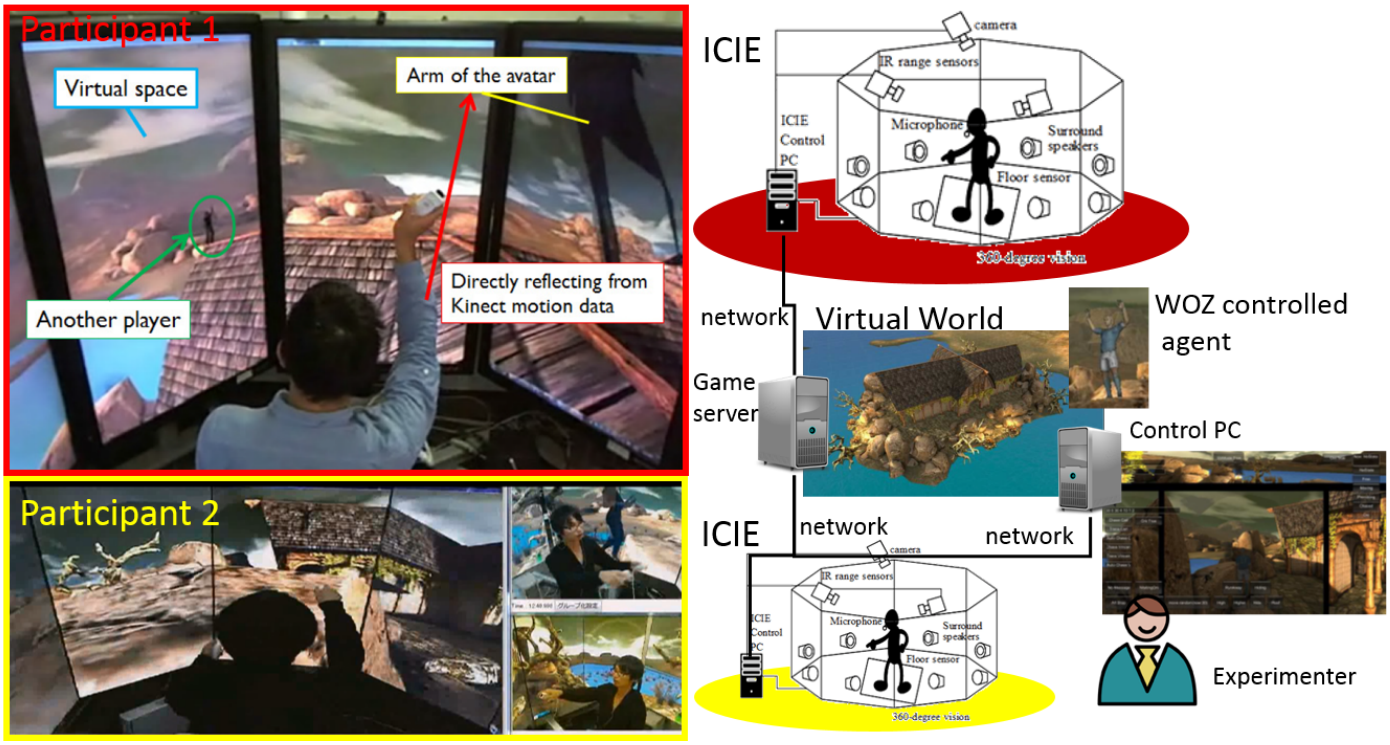


Figure 2. Images during the experiment and the experimental environment.

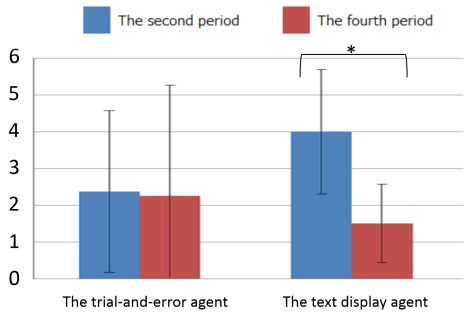


Figure 3. Means of the number of communicative actions towards the agent.

Figure 4. Means of the scores of questionnaires.

though no participant increased in the text display agent group. This means that the trial-and-error agent cannot quickly induce but relatively maintain participant's intentional stance.

2) Questionnaire analysis: The purpose of this analysis is to investigate how the presentation method influenced participants' subjective impressions. The participants answered five rating questions on the ECA's behavior using a seven-point scale. The scale was presented as seven ticks on a black line without numbers, which we scored from 1 to 7. The results are shown in Figure 4. We performed a Mann-Whitney U test on the data in the questionnaire. This analysis shows the final impressions of the agent throughout the experiment.

How human-like do you feel that the agent is?
The average score of the trial-and-error agent was higher than that of the text display agent but there was no significant difference. One reason is that the communication ability of both agents was the same and was poorer than a human's. Since, however, the participants' behavior was changed, we suggest that they unconsciously took the intentional stance.

How strongly do you think the agent has a definite goal?
The participants felt that the trial-and-error agent had significantly more definite goals than the text display agent ($p = 0.039$). This suggests that the agent can effectively provide its goal by performing goal-oriented actions using multimodal behavior. One reason why the text display agent could not do that is that the participants took the design stance because of the artificial "text" and precisely estimated the behavior model.

How smart is the agent?

 ISBN: 978-1-61208-390-2

The participants felt that the trial-and-error agent was significantly smarter than the text display agent ($p = 0.015$). This result also shows that obviously presenting the goal or the intentions is not an effective way to induce the intentional stance.

How well does the agent understand your intentions?

The participants felt that the trial-and-error agent understood their intentions better than the text display agent ($p = 0.054$, marginally significant difference) but the average scores were not high. This result shows the same tendency as the question "how human-like do you feel that the agent is?" The main reason is that the communication ability of both agents was poor. The reason why the text display agent had a lower rating is that the text above the agent's head did not change when the participants communicated.

How much did you enjoy the game?

The participants enjoyed the game significantly more with the trial-and-error agent than with the text display agent ($p = 0.025$), and the scores for both agent were fairly high. This means that the participants were involved in the experiment. We assume that the significant difference was caused by inducing the intentional stance.

V. Discussion

To sum up the experimental results: the trial-and-error agent could not quickly induce participant's intentional stance but, when induced once, the agent maintained intentional stance more than the text display agent. We suggest that the process of constructing a behavior estimation model influences the mental stance participants take to interact with the agent. In addition, an obvious presentation of the inner state of the agent is not effective because the way that an agent presents that is different from the way that a human does.

Clark [10] said that a conversation is a form of joint action. Joint action involves individuals performing individual actions that are intended to carry out a jointly intended shared action. We have also previously proposed that, in some cases of decision-making, the decision or intention is extemporarily shaped, based on the underlying and ambiguous wish (which is one of the sources of the decision and intention) through the interaction [11]. If a person could completely predict and understand a communication partner's behavior and intentions in communication, the communication is the same as conducting a monolog. Therefore, for the text display agent in this study, the participants did not take the intentional stance because they could easily understand the agent's goal. On the other hand, it is difficult to directly understand the goal of the trial-and-error agent from its multimodal behavior. Since the trial-and-error process presented the goal indirectly, the participants had to estimate the goal of the agent through interacting with it.

In future work, we intend to induce the intentional stance and clearly present the goal and the intentions at the same time. We think that a method of implicitly presenting the inner state of agents will be useful for this research.

VI. Conclusions

In this study, we investigated how to induce the intentional stance in human-agent interaction. For this purpose, we tried to induce the intentional stance by presenting goal-oriented behavior in long-term interaction. We proposed two methods of presenting goal-oriented actions and implemented two agents: one was a "trial-and-error agent" that performed goal-oriented actions using multimodal behavior and the other was a "text display agent" that displayed its behavioral intentions via text. We conducted an experiment to evaluate the effect of inducing the intentional stance using these agents. The results showed that participants continuously tried to communicate with the trial-and-error agent, which did not respond to the participant's actions except when necessary for performing the task, and we found that the participants felt that this agent was more goal-oriented, more smart and understood the participants' intentions more than the text-display agent.

References

[1] M. Roubroeks, J. Ham, and C. Midden, "When artificial social agents try to persuade people: The role of social agency on the occurrence of psychological reactance," International Journal of Social Robotics, vol. 3, no. 2, 2011, pp. 155–165.

[2] D. C. Dennett, The intentional stance. MIT press, 1989.

[3] F. Heider and M. Simmel, "An experimental study of apparent behavior," The American Journal of Psychology, 1944, pp. 243–259.

[4] W. H. Dittrich and S. E. Lea, "Visual perception of intentional motion," PERCEPTION-LONDON-, vol. 23, 1994, pp. 253–253.

[5] B. Friedman, P. H. Kahn Jr, and J. Hagman, "Hardware companions?: What online aibo discussion forums reveal about the human-robotic relationship," in Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, 2003, pp. 273–280.

[6] T. L. Chen, C.-H. A. King, A. L. Thomaz, and C. C. Kemp, "An investigation of responses to robot-initiated touch in a nursing context," International Journal of Social Robotics, vol. 6, no. 1, 2014, pp. 141–161.

[7] K. Collins, K. Kanev, and B. Kapralos, "Using games as a method of evaluation of usability and user experience in human-computer interaction design," in Proceedings of the 13th International Conference on Humans and Computers. University of Aizu Press, 2010, pp. 5–10.

[8] Y. Ohmoto, D. Lala, H. Saiga, H. Ohashi, S. Mori, K. Sakamoto, K. Kinoshita, and T. Nishida, "Design of immersive environment for social interaction based on socio-spatial information and the applications." J. Inf. Sci. Eng., vol. 29, no. 4, 2013, pp. 663–679.

[9] "Unity," http://unity3d.com/ (2014/12/01).

[10] H. H. Clark, Using language. Cambridge University Press, 1996.

[11] Y. Ohmoto, M. Kataoka, and T. Nishida, "The effect of convergent interaction using subjective opinions in the decision-making process," in Proc. the 36th Annual Conference of the Cognitive Science Society, 2014, pp. 2711–2716.

 ISBN: 978-1-61208-390-2

Comparing Apples and Orange Cottages

Classifications and Properties

Julia M. Taylor
Computer and Information Technology & CERIAS
Purdue University
West Lafayette, Indiana, USA
jtaylor1@purdue.edu

Victor Raskin
Linguistics & CERIAS
Purdue University
West Lafayette, Indiana, USA
vraskin@purdue.edu

***Abstract*—This paper deals with the rules of good classification and comparison, as well as matching the representation of the results with what has actually been accomplished. The emphasis in machine learning classifications, as well as, sometimes, outside of that paradigm, is almost exclusively on the precision of separating classes from each other, and hardly any effort is made to assess the nature of the classes with regard to their grain size. This results in a considerable disparity between the claimed results and what is really demonstrated, leading in turn to crude solutions to issues and poorly functioning applications. We propose an ontological solution, following the explicit tracing of a conceptual hierarchy underlying the classes. This approach may lead to a variety of solutions that can be compared after classification and similarity studies mature enough to face the issue.**

Keywords-comparison; classification; hierarchy; propert;, similarity; ontology; concept.

I. INTRODUCTION: SLOPPY CLASSIFICATIONS

Research, cognition, reasoning all involve some comparison, classification, similarity. Decisions on the bi-, tri- or multifurcation of a concept are common and inevitable. Statistical methods discover and refine unknown classifiers to divide a large bunch of samples into in- and outliers. Rule-based systems use rules to compare and classify. Are we doing it right? Do we know how to do it right? Is it useful to do it right?

The problem addressed in this paper is the status and (mis)interpretation of classifications and classifiers. Do the researchers have a clear picture of what they actually compare as opposed to what they want or purport to compare?. It addresses primarily the lack of attention and direct research effort in clarifying and codifying this problem—actually, an amazing lack of awareness that exists. Yet, grain size misclassification can have devastating effects on understanding the phenomena and question and the issues with them, as well as recommending precise solutions, pretty universally across research, from political and military solutions to treating bad cells in a patient.

The research question, then, that we are posing here, quite possibly for the first time so explicitly, is how to clarify and raise the precision of a proposed classification in just about any area of research. We will propose that the solution requires an ontological framework and a clear notion of grain size. The paper is not meant as a critique of the status quo with regard to the treatment of classification but rather to inform the diverse communities of scholars of a primising framework for improving that treatment. But first, a couple of examples of classification inexactitude, with consequences. Both come from areas of research where extensive scholarship has been done, including our own, but other than that, all they have in common is something they share with virtually any other area of research, namely, that they do classification and interpret their results.

Recently, we were asked to review a paper (not yet published, it may become officially citable soon) that used a machine-learning approach to separate serious text from satirical one. Not surprisingly, the results were statistically significant. Moreover, one could look at the features that were used to make the classification. A question to be asked is whether one should look at such features to shed light on the properties of satire. A simplified schema of humor classification in Figure 1 helps to see why it is not desirable (see Raskin et al. [1] for a discussion of the state of the art in humor theory and computational humor and for multiple references; cf. Raskin [2])..

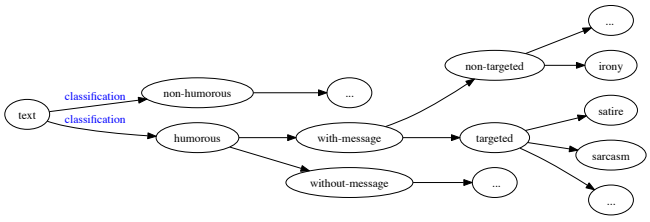


Figure 1. Simplified humor classification.

The question to be asked when one looks at the features is: was the distinction that was caught really the one between non-humorous text and satire? From the figure above, which most humor researchers will consider simplistic perhaps but plausible, satire is not just humorous but also containing a message that is targeted. It is also distinct somehow from irony and sarcasm. None of these features is mirrored on the non-humorous side, and there is a very serious risk of misinterpreting the results of the experiments. In all likelihood, what the classification captures is the distinction, at a much coarser grain size, between humorous and non-humorous text, the latter being a remote ancestor

 ISBN: 978-1-61208-390-2

(hyperonym) of satire and impervious to the targeted-message nature of satire.

Similarly, in our own recent work on phishing (see Park et al. [3], Stuart et al. [4], and Park and Taylor [5], we compared *bona-fide*, legitimate email in the Enron corpus with a bunch of known phishing emails. We invested a considerable theoretical and methodological effort in the work and got reportable results. But, we dealt with a situation that is similar to satire, with phishing being the counterpart of satire (see Figure 2).

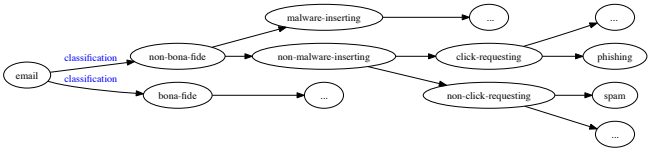


Figure 2. Simplified email classification.

The notions in Figure 2 are better definable than those in Figure 1; so in order to identify phishing, we needed first to have a corpus containing both bona-fide email from non-bona-fide emails. After separating those out, we needed to focus on the non-bona-fide corpus and separate malware-inserting emails from non-inserting; and then on, to a still narrower corpus and separating click requesting from non-click requesting. If the only click-requesting kind is phishing we are home, right? No, actually, there still exists *bona-fide* email that is click-requesting because our graph is actually not a tree but rather a lattice. In any case, we did not provide for any such complications, so we did just the first separation, and we failed to separate phishing from any other kind and sub-kind of non-bona-fide email. Our excuse, if any, is that we did not have enough corpora for the lower divisions. It is the same excuse as in the case of satire above. One does depend heavily on the availability of sufficient corpora but this is not a sufficient excuse for misidentifying the results.

The two examples above are sloppy classifications, and those are ubiquitous. Section II seeks help from adjacent disciplines, namely, the philosophy of science for theory building and research hygiene: as well as from psychology for similarity studies. Section III introduces the Ontological Semantic Technology (OST) whose property-rich ontology is a suitable base for rendering classification more rigorous and precise. We believe that more approaches will be developed to handle various meaning-based data- and text-processing applications, and that will be the time to compare OST to competition. We are not sure, however, that without a similar proper ontological base, a solution is possible. Section IV formalizes the OST approach, with a focus on ontological concepts and properties. The conclusion of Section V puts forth the down-to-earth application of the principle of rigorous comparison and application: where sole-property comparison is impossible or impractical, just explicating the property-set comparison may be a path to success. Given the paucity of effort in ensuring the grain size rigor of classifications and comparisons, the main contribution of this paper is drawing the wide community's attention to the issue of sloppy classifications, especially when the features are used to understand the nature of the crucial role of ontologies as remedy.

II. STATE OF THE ART

All research requires definitions, distinctions, comparisons, and classifications. The need to introduce categories and sub-categories is universal. Surprisingly, the state of the art on the precision of classification is minimal: there is no precision metric nor evaluation procedure for doing it right, and there is a definite, if not desperate need in both for virtually any area of research. In this section, we overview research on the philosophy of science that is supposed to contribute to theory building and psychology, mostly, cognitive psychology, on similarity (see references below). We briefly look for help in heuristics as well

A. Philosophy of Science

A brief look at the index pages of a couple of new readers in the field (Curd and Cover [6], Balashov and Rosenberg [7]) discovers a shared lack of any mentions of classification, comparison, distinction, separation, or hierarchy as worthy items of discussion. Hardly anything comes up on web and library searches. The last item does emerge in the context of biological classifications, and this should be expected: Linnaean classifications of the animal world, yet another unrelated domain of classifying, along with humor research and phishing mentioned above, have traditionally provided poster-child examples of straightforward sub-classifications, such as shown in Figure 3:

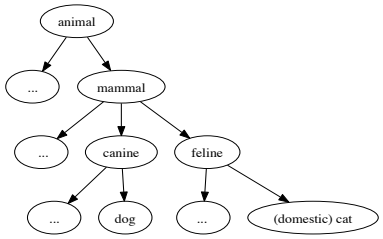


Figure 3. Simplified animal classification.

What Ereshefsky [8] states on the very inside cover, however, is as follows:

> "The question of whether biologists should continue to use the Linnaean hierarchy is a hotly debated issue.Invented before the introduction of evolutionary theory, Linnaeus's system of classifying organisms is based on outdated theoretical assumptions and is thought to be unable to provide accurate biological classifications.
>
> Marc Ereshefsky argues that biologists should abandon the Linnaean system and adopt an alternative that is more in line with evolutionary theory."

The customary advantage of the ancient classification is what was introduced and studied in the 20th-century mathematics as inheritance (Touretzky [9]): mammals inherit all properties of animals and add a few extra properties of their own; canines and felines inherit all of those, and each adds an extra set of additional properties; cats and dogs add another set of properties as well. As a result, a dog collects

 ISBN: 978-1-61208-390-2

all the properties from animal (and its superclass, if any) to canine, as well as adding its own properties that other canines do not have.

Perhaps, one explanation of classifications and hierarchies not being actively discussed and researched is that, as per Potochnik and McGill [10],

> "The concept of hierarchical organization is commonplace in science and philosophical treatments of science. Though there are different applications of the concept of hierarchy, our primary focus here is the idea that material composition is hierarchical. Subatomic particles compose atoms, which compose molecules; cells compose tissues, which compose organs, which compose organisms; interbreeding organisms compose populations, which compose communities, which compose ecosystems; and so on. The basic idea is that higher-level entities are composed of (and only of) lower-level entities, but the prevalent concept of hierarchical organization involves stronger claims as well. The compositional hierarchy is often taken to involve stratification into discrete and universal levels of organization. It is also often assumed that levels are nested, that is, that an entity at any level is composed of aggregated entities at the next lower level."

The few references that are there to classifications, hierarchies, and levels in the contemporary philosophy of science seem to be all derived from an almost forgotten classic [Feibleman {11], p. 59], where the very first of the many rules establishing the hierarchy of "integrated levels" states that

> "[e]ach level organises the level or levels below it plus one emergent quality. Thus the integrative levels are cumulative upward. This proposition implies that everything has at least the physical properties and has led to the position of supreme importance of the physical world in science and philosophy."

Very characteristically to this strand, the whole philosophy of levels and hierarchies is limited to the physical world: the last sentence of the quote above limits it to physical objects, typically starting from bottom up with atoms and molecules. Potochnik and McGill [10] follows the same route, even though the paper applies this philosophy to ecology. Ereshefsky [8] is all about biology. So, Attardo and Raskin [12], an additional useful source on humor theory, had to do its own philosophy of science when it needed to establish a hierarchy of abstract levels of representation for a verbal joke in the General Theory of Verbal Humor on the principle of each higher level adding a restriction on the lower level, thus narrowing that latter's scope of included phenomena, as per Figure 4.

The integrative levels theory was, apparently, running high and ambitious in the mid-third of last century (see Bertalanffy and Woodger [13], Novikoff [14], and Bertalanffy [15]0, prompting Feibleman to hope, after Bertalanffy [15], for "a sort of super-science which shall have as its subject-matter the relations between the sciences. The philosophy of science may yet be the source for the development of an empirical field itself consisting of the integrative levels, a sort of meta-empirical field, with its own entities and processes and laws" (Feibleman [11], p. 59]. This has never happened, and this paper is suffering from the lack of helpful pertinent wisdom.

Figure 4. Levels of hierarchy of GTHV.

B. *Formal Ontology*

Some of what the philosophy of science could have delivered was contributed in formal ontology (see, for instance, Guarino and Poli [16]): a rigorous notion of hierarchy and inheritance, with a detailed study of subsumption. Never directly allied with the philosophy of science, it has been a blend of philosophy and logic, not focusing on the building nor application of actual ontologies and thus not involving itself in comparisons and classifications—just contributiung to a solid theoretical foundation for doing it right.

C. *Cognitive Psychology*

The main contribution that cognitive psychology has made for classification and hierarchy is a bit indirect: distinction and classification is closely related to the notion of similarity: note that, in any hierarchy, the subclasses of the same class share all the inherited properties and differ only in those extras that they add, and it was a high-powered strand of research on similarity and properties that flourished in cognitive psychology in several previous decades.

It was, apparently, Gregson [17] that put the measurement of (perceived) similarity on the map of psychological research. His thorough survey of similarity models, spatial and otherwise, did not focus, however, on the foundational notion of property that similarity must be based on, and it was the seminal Tversky [18: p. 330] that did. It proposed the general format for a property-based measurement function of similarity as "s(a,b) = F(A∩B, A - B, B - A)," where "[t]he similarity of a to b is expressed as a function F of three arguments: A∩B, the features that are common to both a and b; A — B, the features that belong to a but not to b; B — A, the features that belong to b but not to a."

In subsequent usage, the formula above has been often traded for a cruder but simpler measure as A∩B/A∪B, i.e., the intersection of the feature sets of a and b, divided by the union of these sets, standardly normalized to the [0,1] interval.

It is, however, Osherson et al. [19][20] that built a series of similarity models on a couple of subsets of a small animal dataset underlaid by a somewhat greater set of their properties, 48 and 85, respectfully for the latter paper. First having calculated the similarity measurements among the animals from the data set, using the simplified metric above,

 ISBN: 978-1-61208-390-2

they conducted an experiment with 10 human subjects and compared their similarity judgments with those predicted by their model.

There are two aspects of this research that are of a particular interest to us here. First, the origin of the properties used as the foundation of the similarity model: they were compiled by the researchers "empirically" and offered to the 10 subjects with the instruction to detect and suggest the addition of any new property other than the sounds the animals made. An additional property would only be taken into consideration if proposed by more than one subject. We proposed an ontological foundation for our (Taylor and Raskin[19]).

Second, none of Osherson's and his associates' papers over almost two decades, directly based on the animal dataset or following from earlier research on it, had the similarity model as the goal. In fact, the models were obtained to be used as tools in research on human reasoning, such as inductive judgments (Osherson et al. [19]), default probabilities (Osherson et al. [20]), the conjunction fallacy (Tentori et al. [22]). Later related work (Perfors et al. [23], Kemp et al. [24]), using their own variations of animal datasets and properties, applied their models to research the way children learn "domains" and "theories of the world," respectively. We are also interested not so much in similarity models as in the nature of properties that are out there in the world and that people reason with, thus necessitating the need to computerize those properties for the purpose of constructing a meaning-based structure from text and other data.

D. Heuristics

Our best help should have probably come from this step-daughter (hopefully, Cinderella) of mathematics, pretty much completely ignored by other disciplines. The insights in the old classic Polya [25] and the newer classic Pearl [26] should inform theory-building significantly. In fact, heuristics should be the basis of any graduate course or seminar on research methods on top, if not even instead, pure statistics that most universities offer exclusively. Unfortunately, heuristic ideas are hard to pack in off-the-shelf software, and abduction, on which much heuristics rests, has not been able to compete with deduction and induction, instead of its ubiquity and scope, for the minds of scientists and other scholars.

III. ONTOLOGY, HIERARCHY, AND GOOD CLASSIFICATION

A. Ontology

The ontology comes from our particular approach to computational semantics the Ontological Semantic Technology (OST). The theory-cum-technology is a radical revision and improvement of Nirenburg and Raskin [27];see Raskin et al. [28], Taylor et al. [29], Taylor and Raskin [30], Hempelmann et al. [31], Taylor et al. [32][33]. The centerpiece is indeed the language-independent semi-automatically constructed engineering ontology, as per Gruber [34], consisting of concepts (OBJECTS and EVENTS) linked with a rich system of PROPERTYS. Each supported natural language, e.g., English and Russian shown on Figure 5, has a lexicon with supporting morphological, and syntactic rules, supplemented with phonological rules (not shown), when required by an optional speech recognition functionality.

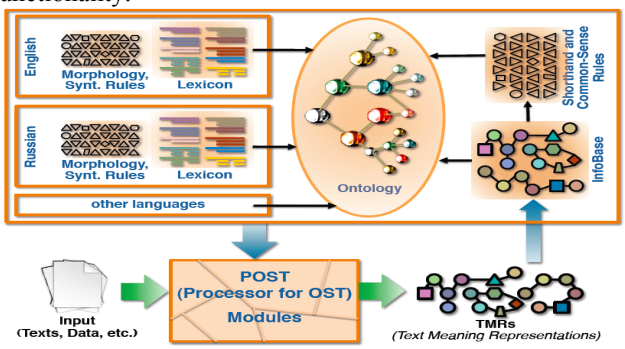


Figure 5. OST Architecture.

A lexicon contains all the lexical entries for the language, each entry with all of its senses. The central components in an entry are the partial syntactic information (SYN-STRUC) and full semantic information (SEM-STRUC). The latter typically "anchors" the sense of a lexical item in the appropriate ontological concept, restricts some of its property fillers if necessary, and binds the variables introduced in the sense's SYN-STRUC. The ontology captures much information about how things are in the single or multi-domain world it serves. The ontology is supplemented by the previously processed information from the InfoBase and by dynamically collected shortcut and commonsense rules on which the machine and a human ontology engineer collaborate.

The OST various processors operate on the anchoring results. When a sentence arrives at its input, the analyzer looks up every word in the lexicon, checks that its usage conforms to the SYN-STRUC, selects the corresponding SEM-STRUC, identifies the event in each clause, and then attempts to match all the other concepts evoked by the words and phrasals as the fillers of the event's properties. The result, the Text Meaning Representation (TMR) of the input sentence is stored in the InfoBase for further usage, including possible correction or challenge by the later arriving text.

B. Hierarchy

The OST ontology, like most real and pseudo-ontologies, is based on subsumption, which means that its IS-A property is privileged to pass on properties from higher-level (parent) nodes to low-level (child) nodes, as shown in Figure 3. The ontology is not a tree, however: rather, it is a lattice. This means occasional complications to inheritance. Also, very rarely, a property to be inherited have to be blocked as, for instance, continuing with the convenient animal world, ostriches and chickens should inherit all properties of birds except the ability to fly.

C. Good Classification

A good classification is a minimal, carefully controlled deviation from the ideal classification, a deviation which occurs only when necessary. The ideal classification is

 ISBN: 978-1-61208-390-2

achieved by a hierarchy, in which each child adds one simple ontological property to the ones it inherits from its parent and, thus, from its entire ancestry. In reality, what is added is most often a set of properties. An ontological approach allows us to be fully aware of what the set consists of and, if necessary, to entice us to separate its property elements out in additional computer experiments. Our ongoing work on composable properties (Taylor and Raskin [35]) promises further progress in this direction.

IV. A BIT OF FORMALISM

For the purposes of this discussion, let us assume that each concept C can be defined as a combination of properties P_1, P_2, …, P_n. Each of these properties can be further restricted by specifying a particular argument to a property. For example, a concept HUMAN (see Figure 6) can have a property GENDER, and for the sake of simplicity, let us assume that the range of this property can be either MALE or FEMALE. In order to define the concept WOMAN, one would need to restrict the property GENDER from its most generic case to that of only FEMALE.

Now, let us consider a more general case. Suppose a concept C1 can have a property P1. Each of its children, C2.i will also have a property P1, restricted by a particular filler -- let us say filler a_i -- as well as inheriting the rest of the properties that C1 inherited, each with the restrictions done for the parent (see Figure 7). Children of C2.i will also have some property, lets call it P2.i, that will be restricted by some fillers, thus introducing a new layer of descendent concepts, all of which will still inherit P1(a_i). Eventually, we will run into a situation where a concept Cj,k is composed by j-1 properties, each of which is restricted by at least once when passed from a ancestor to a descendant. It is possible that concept Cn.i (see Figure 7), which inherited a property P2.3 with filler a8 needs to have a more specific filler than that of a8. Then, the property Pn.i will be the same as property P2.3, but the filler of the property for the children of Cn.i will be have to be children of a8.

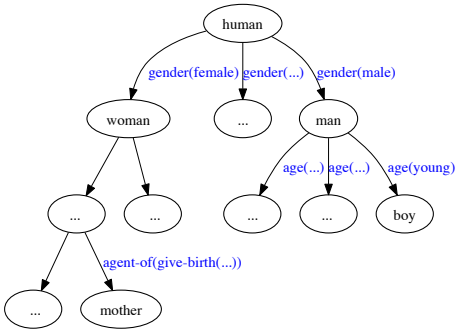


Figure 6. Hierarchy example.

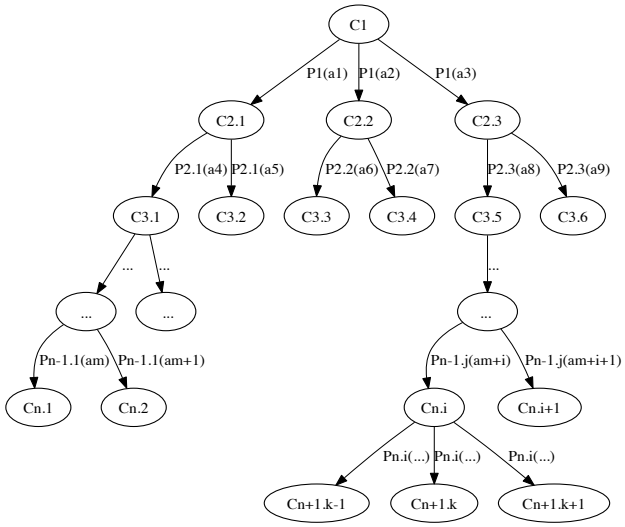


Figure 7. Hierarchy based on properties.

Within such a structure, the concept C3.1 differs from the concept C3.5 by a filler of property P1, as well as by properties uniquely introduced for these concepts, namely, P2.1 and P2.3. Perhaps, it will be clearer in a more concrete example, outlined in Figure 6. The concept MOTHER differs from the concept BOY by a filler of the property GENDER, as well as by the properties that lead to the concept MOTHER, and those that lead to concept BOY. In Figure 6, the only ones that are visible are AGE and AGENT-OF. It is possible that some of these properties are the same – we can use our common sense here—both MOTHER and BOY do have some AGE, and both are likely to be AGENT-OF something. In that case, we will say that property Px and Py are the same.

It is also possible that P1 is entirely inherited by the child concept, without any restriction, but then there will be another property Pi that the child will restrict, otherwise the child and parent concept run the danger of being identical.

If we are not dealing with a tree, but with a lattice, there is more flexibility in building a graph that would absorb the common properties within their parents, whether it would result in multiple parenting scenarios or not. There are many formalisms that allow such lattice to be build and construct it automatically.

Two concepts Ci and Cj, then, can be looked at in terms of 3 sets of properties: one set that is shared between them, one set that Ci has, but not Cj and one set that Cj has but not Ci. To complicate matters, the shared set is likely to have different fillers of the properties, and thus we have to subdivide this set into properties that have identical fillers between Cj and Cj and properties that have different fillers.

 ISBN: 978-1-61208-390-2

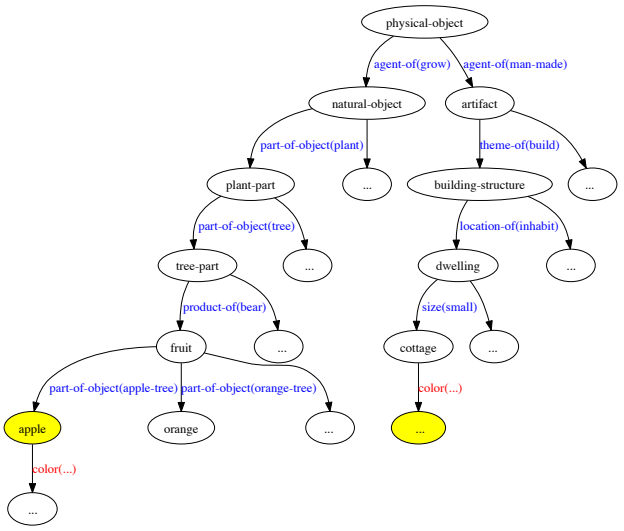


Figure 8. A sample structure to lead to apple and cottage concepts.

As an example, lets us compare apples and orange cottages, as highlighted in Figure 8. The overlapping property set between orange cottage and apple (also orange?) is {AGENT-OF, COLOR}. The other two sets are {PART-OF-OBJECT, PRODUCT-OF} and {THEME-OF, LOCATION-OF, SIZE}. If we are not paying attention, it could be said that a differentiating feature between apples and orange cottages is the fact that one of them is part of a tree, and the other one is not. While that is definitely true, any dwelling is not part of a tree, and neither is any artifact. This happens because the PART-OF-OBJECT property is in the set of properties that only one of the compared concepts has. On the other hand, we could compare orange cottages and apples on the property COLOR (e.g., the color of this apple is exactly like the color of my cottage).

Notice also that the closer the concepts are in their hierarchy, the easier it is to compare them – because, again, the overlapping property set is large. Thus, we can compare apples and oranges on many more parameters than apples and cottages (orange or not). Of course, comparing apples and apples is even better – we can just count the properties in the overlapping set.

If we can rely on such sets of properties and hierarchies, it is easy to see why a whale can be both compared to mammals and fish, even if we (as young children?) may not know where exactly it belongs in the hierarchy.

V. CONCLUSION: CUI BONO?

In this paper, we discussed the hygiene of good classification and comparison and suggested that an ontological foundation would considerably clarify the issue. Even if comparison on one sole property is not attainable or even desirable, and subclasses have to differ from their ontological parents by a set of features, it is useful to be fully aware of it and be prepared to un-bunch them if necessary. It is also crucially important not to misstate nor to misrepresent the result by a hierarchical confusion. Let us not compare apples with orange cottages without figuring out explicitly what properties separate them and their grain size correspondence.

REFERENCES

[1] V. Raskin and J. M. Taylor, “On the tansdisciplinary field of humor research,” Journal of Integrated Design and Process Science 16(3), pp. 133-148, 2013

[2] V. Raskin (ed.), The Primer of Humor Research, Berlin: Mouton de Gruyter, 2008.

[3] G. Park and J. M. Taylor, “Towards text-based phishing detection,” Proc. SDPS-13. Sao Paolo, Brazil, 2013.

[4] L. M. Stuart, G. Park, J. M. Taylor, and V. Raskin, “On identifying phishing emails: Uncertainty in machine and human judgment,” Proc. NAFIPS-14, Boston, MA, 2014.

[5] G. Park, L. M. Stuart, J. M. Taylor, and V. Raskin, “Comparing machine and human ability to detect phishing emails,” Proc. IEEE-SMC-14. San Diego, CA, 2014.

[6] M. Curd, J. A, Cover, and C, Pincock, eds., Philosophy of Science: The Central Issies, 2nd ed. New York: Norton, 2013.

[7] Y. Balashov and A. Rosenberg, eds., Philosophy of Science: Contemporary Readings. London-New York, Routledge, 2002.

[8] M. Ereshefsky, The Poverty of the Linnaean Hierarchy: A Philosophical Study of Biological Taxonomy. Cambridge, UK: Cambridge University Press, 2001.

[9] D. S. Touretzky, The Mathematics of Inheritance Systems. Los Altos, CA: Morgan Kauffman, 1996.

[10] A. Potochnik and Brian McGill, “The limitations of hierarchical organization,” Philosophy of Science 79:1, 2012.

[11] J. K. Feibleman, “Theory of integrative levels,” British Journal for the Philosophy of Science 5 (17), 1954, pp. 59-66.

[12] S. Attardo and V. Raskin, “Script theory revis(it)ed: Joke similarity and joke representation model,” HUMOR: International Journal of Humor Research 4:3-4, 1991.

[13] L. von Bertalanffy and J. H. Woodger, Modern Theories of Development. Oxford: Oxford University Press, 1933.

[14] A. B. Novikoff, “The concept of integrative levels and biology, “ Science 101, 1945, pp. 209-215.

[15] L. von Bertalanffy, “An outline of general system theory,” The British Journal for the Philosophy of Science I-134, 1950.

[16] N. Guarino and R. Poli, Eds., Special Issue: The Role of Formal Ontology in the Information Technology, International Journal of Human and Computer Studies 43:5-6, 1995.

[17] R. Gregson, Psychometrics of Similarity. New York: Academic Press, 1975.

[18] A. Tversky, “Features of similarity,” Psychological Review 84: 4, 1977, pp. 327-352.

[19] D. N. Osherson, O. Wilkie, E. E. Smith, A. Lopez, and E. Shafir, “Category-based induction,” Psychological Review 97:2, 1990, pp. 185-200.

[20] D. N. Osherson, J. Stern, O. Wilkie, N. Stob, and E. E. Smith, “Default probability,” Cognitive Science 15, 1991, pp. 251-269.

[21] J. M. Taylor and V. Raskin, “Understanding and structuring language descriptions: The case of 101 animals,” Proc. IEEE SMC 2012. Seoul, S. Korea, 2012.

[22] K. Tentori, N. Bonini, and D. Osherson, “The conjunction fallacy: A misunderstanding about conjunction?” Cognitive Science 28, 1990, pp. 467-477.

[23] A. Perfors, C. Kemp, and J. B. Tenenbaum, “Modeling the acquisition of domain structure and feature understanding, “ Proc. CogSci-05, 2005.

 ISBN: 978-1-61208-390-2

[24] C. Kemp, J. B. Tenenbaum, T. L. Griffiths, T. Yamada, N Ueda, "Learning systems of concepts with an infinite relational model," Proc. AAAI-06, 2006.

[25] J. Polya, How to Solve it: A New Aspect of Mathematical Method. Princeton, NJ: Princeton University Press, 1945.

[26] J. Pearl, Heuristics: Intelligent Search Strategies for Computer Problem Solving. New York: Addison-Wesley, 1984.

[27] S. Nirenburg and V. Raskin, Ontological Semantics. Dordrecht: D. Reidel, 2004.

[28] V. Raskin, C. F. Hempelman, J. M. Taylor, "Guessing vs. knowing: The two approaches to semantics in natural language processing,: Proc. Dialogue 2010. Bekasovo/Moscow, Russia, 2010, pp. 642-650.

[29] J. M. Taylor, C. F. Hempelmann, V. Raskin, "On an automatic acquisition toolbox for ontologies and lexicons in ontological semantics," Proc. ICAI-10, Las Vegas, NE, 2010, pp. 863-869.

[30] J. M. Taylor and V. Raskin, Understanding the unknown: Unattested inpiut in natural language," Proc. FUZZ-IEEE-11. Taipei, Taiwan, 2011.

[31] C. F. Hempelmann, J. M. Taylor, V. Raskin, "Application-guided ontological engineering, " Proc. ICAI-10. Las Vegas, NE, 2010.

[32] J. M. Taylor, V. Raskin, C. F. Hempelmann, "From disambiguation failures to common-sense knowledge acquisition: A day in the life of an ontological semantic system," Proc. WI-IAT-11. Lyon, France, 2011.

[33] J. M. Taylor, V. Raskin, C. F. Hempelmann, "Towards computational guessing of unknown word-meanings: The ontological semantic approach," Proc. CogSci-11. Boston, MA, 2011.

[34] T. R. Gruber, "Toward principles for the design of ontologies used for knowledge sharing," in: [14], 1995, pp. 907-928.

[35] J. M. Taylor and V. Raskin, "On the nature of composable properties," Proc. CCAHI Workshop, AAAI-14, Quebec City, Canada 2014.

 ISBN: 978-1-61208-390-2

Towards Identifying Ontological Semantic Defaults with Big Data: Preliminary Results

Tatiana Ringenberg, Julia Taylor, John Springer
Computer & Information Technology
Purdue University
West Lafayette, IN, USA
{tringenb, jtaylor1, ja}@purdue.edu

Victor Raskin
Linguistics & CERIAS
Purdue University
West Lafayette, IN, USA
vraskin@purdue.edu

***Abstract*—This paper reports on a work-in-progress and suggests a method of detecting conceptual defaults in natural language big data. It combines Hadoop and Nutch technologies for web crawling with the Ontological Semantic Technology (OST) in an initial effort of this kind. Initial results demonstrate the viability of this method to detect unintended inference within text.**

Keywords-big data, Hadoop, Nutch, conceptual default, Ontological Semantic Technology.

I. INTRODUCTION

This paper merges big data research technology with the important computational semantic task of identifying conceptual defaults, i.e., the parts of text that the speaker/writer omits because they are too obvious to mention, both for himself/herself and for their intended audience—and occasionally for all. Thus, hardly anybody would say, *I unlocked the door with the key*, preferring to drop the prepositional phrase as obvious [1]-[3]. The prepositional phrase is, then, a conceptual default. On the other hand, if a competent speaker does verbalize a default, the hearers may suppose that the prepositional phrase is not the default: for instance, if all the locks in the building are electronic. The defaults are very important for the computer to be aware of and use for inferences and reasoning as people do.

Though little to no work has been done on defaults outside of Ontological Semantics Technology (OST), defaults can be loosely related to Grice's Maxim of Quantity which states that a person will not mention more than what is necessary in a conversation [11]. This definition is quite broad. The work in this research seeks to solve the problem of identifying a small portion of information which a speaker considers to be too trivial to mention.

The algorithm used in this research is intended to pull only unintentional inference related to verb events, nouns and adjectival modifiers. In previous work, this specific type of default has been referred to as White Dude Inference (WD-Inference) [2].

Big Data is used in this research as a method of confirmation of our implementation of the WD-Inference algorithm. The website purdue.edu was chosen due to the abundance of texts. The website also fits very well with the principles of Big Data including high volume, velocity and variety of information assets [13]. As such the data is prevalent, and varied, enough to test the algorithm.

This paper describes the implementation decisions and steps in implementing the algorithm for basic WD-Inference-style defaults. Section II describes the background information related to Ontological Semantics, Ontological Semantics Technology and Big Data. Section III describes the materials used for implementation of the algorithm. Section IV describes the methods used in the preliminary study. Section V describes preliminary results of the research. Section VI describes the next steps that will be taken to improve and complete the research. Section VII describes future work related to Ontological Semantics defaults and OST.

II. BACKGROUND

A. Conceptual Defaults

Surprisingly little has been done about conceptual defaults before [1]-[3]. These papers could not have emerged without the Ontological Semantic Technology [4]-[9] that emerged from Ontological Semantics of the 1990s [10], and made it possible to develop the comprehensive human-like meaning text representation of text on the basis of an engineered language-independent ontology and a set of language-specific lexicons, whose every sense of every entry is anchored in ontological concepts linked with multiple properties. A pattern-matching and a graph-producing analyzer, the central POST elements, compete with each other in Text Meaning Representation (TMR) production. Figure 1 shows the OST architecture.

It would be a stretch to consider Grice's [11] Maxim of Quantity or scarce work on semantic ellipsis [12] as bearing on defaults but it may provide some comfort to a novice because these sources do bear somewhat on what is not necessary to say and how the text that is there may help to reconstruct the text that was elided, in some cases, because it was obvious.

To the knowledge of the researchers, no computational implementation of defaults has previously been created either. As such, there is no significant work to which we may compare the performance or the accuracy of the proposed solution. The solution provided in this research is intended to be a first step towards automatic acquisition of Ontological Semantics defaults.

 ISBN: 978-1-61208-390-2

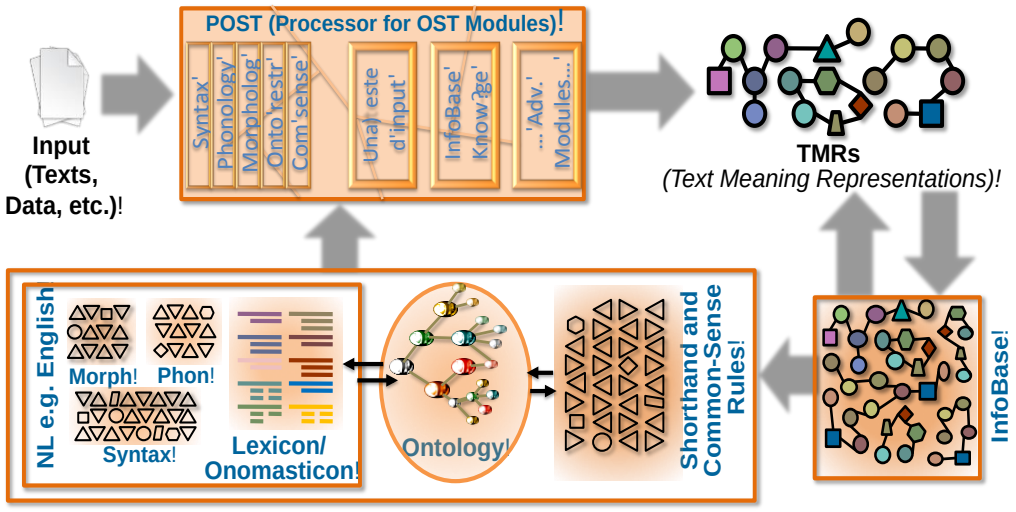


Figure 1. OST Architecture

B. Big Data

According to Gartner, Inc. [13], "Big data in general is defined as high volume, velocity and variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making." As depicted in Figure 2, at the heart of Big Data is analysis and refinement leading to more effective decision-making.

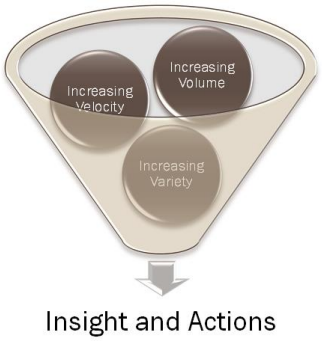


Figure 2. Big Data's 3V Leading to Insight and Actions

Moreover, Big Data lies at the confluence of several fields and disciplines – including Computation and Cyberinfrastructure, Visual Analytics and Visualization, Ethics, and Quality Assurance (see Figure 3) – and in its most effective application, has a context in a particular domain such as Business/Finance, Social Sciences, Life Sciences, Physical Sciences, and Engineering. It is in leveraging the intersection of all of these areas that Big Data delivers its greatest value.

Big Data is also pervasive. According to a report from McKinsey [14], "[l]eaders in every sector will have to grapple with the implications of big data, not just a few data-oriented managers. The increasing volume and detail of information captured by enterprises, the rise of multimedia, social media, and the Internet of Things will fuel exponential growth in data for the foreseeable future."

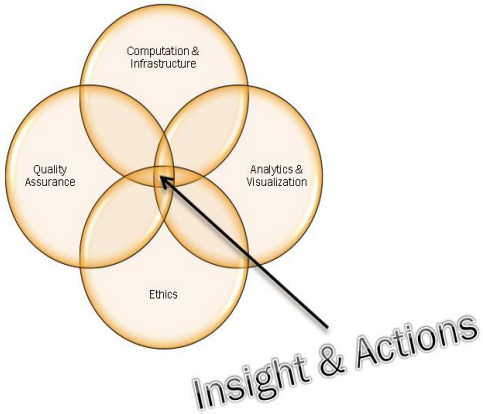


Figure 3. Confluence of Fields and Disciplines

Big Data projects cover a gamut of areas and uses. These include such well-known scientific projects as the Large Hadron Collider and the Large Synoptic Survey Telescope as well as the daily operations of Facebook and Google.

Given its dimensions related to variety and velocity, Big Data has an obvious relationship with Natural Language Processing as natural language represents a frequently generated source of "unstructured" data.

To capitalize on this ideal source of Big Data, we leverage a technology synonymous with Big Data, Hadoop, and one of the solutions for web crawling, Nutch, built on it. By leveraging these tools, we lay the foundation for scalability beyond small data sets for our NLP needs.

C. Purpose

The overall goal of this research is to provide a mechanism for the identification of a very small subset of

 ISBN: 978-1-61208-390-2

defaults. Specifically, this research examines the relationships between events (verbs) and the nouns and adjectives related to them. It is hypothesized that defaults should not show up within text unless modified. This means that the default for *drive* is *car* (the entity that is being driven), *car* should very rarely show up in text with the verb *drive* by itself. The only time it would be acceptable to see *car* with *drive* would be when the speaker does not generally drive a car or when a description of a car is provided. For instance, it is unusual to say *I drive a car*, because that's too trivial, but *I drive a red car* is fine, especially, if this information about a car is new.

In this study, we examine the 200 most common verbs in Brown corpus along with the noun and adjective arguments that accompany them. We compare the list of verbs with no arguments, verbs with a noun argument and verbs with a noun and adjective argument in order to find and analyze potential defaults. Future work will examine other types of defaults.

III. SELECTION OF MATERIALS

A. Parser Tool Selection

In order to pull verbs, adjectives and nouns from verb phrases a parser was required. The goal for these parsers was to allow the researcher to pull adjectives and nouns that modified a particular verb. Stanford Parser 3.4.1 was chosen for its popularity within Computational Linguistics and its parsing flexibility [16].

B. Corpus and Verb Selection

As this research seeks to analyze the relationships between verbs, nouns and adjectives specifically, a tagged corpus was ideal for the preliminary stages of research.

Brown Corpus was specifically chosen for this task because of its size, part-of-speech tagging and wide use within Computational Linguistics. Brown Corpus is a collection of American English documents from the 1960's. It consists of about 500 samples and around a million words [15].

Using Brown Corpus, all verbs were pulled from the sentences and stemmed using Porter Stemmer. Stemmed verbs that the researchers believed would not provide relevant information were removed from the list of verbs. The verbs that were removed include *say, be, go, get, "have", state*, and their forms. Verb frequencies were then calculated and the 200 most frequent verb stems were selected.

C. Structure Selection

Stanford Parser was used to generate typed dependencies. Typed dependencies are used to find relationships between words in a sentence. The goal is to provide syntactically and (partial) semantically useful information about the relationships between words. The following is a dependency for the sentence *Jackie Brandt singled deep into the hole at the short to start the rally*:

nn(Brandt-2, Jackie-1)
nsubj(singled-3, Brandt-2)
nsubj(start-11, Brandt-2)
root(ROOT-0, singled-3)
advmod(singled-3, deep-4)
det(hole-7, the-6)
prep_into(singled-3, hole-7)
prep_at(hole-7, short-9)
aux(start-11, to-10)
xcomp(singled-3, start-11)
det(rally-13, the-12)
dobj(start-11, rally-13)

Dependencies were chosen as the primary tool for sentence analysis due to simplicity. Though Dependency Grammars are not as expressive as syntax trees, they make the relationships between verbs, adjectives and nouns more transparent.

D. Crawl Selection

Brown Corpus consists of documents created in the 1960's. As such, it is the researchers' belief that a more up-to-date corpus is needed to confirm the relevance and accuracy of the methodology described in this paper.

It is also our belief that this methodology must be scalable. Given the large number of blogging and social networking venues, data these days is both prevalent and large. As such, we apply our methodology to a larger dataset than Brown Corpus.

In order to confirm the methodology for extracting defaults, described in this paper, Hadoop and Nutch were used. As was mentioned earlier, Hadoop is often used to work with Big Data and Nutch provides a very easy and intuitive web crawling experience that goes well with Hadoop.

The website "purdue.edu" was used as the initial seed for the crawl. The Purdue website was chosen due to its terms-of-service, the vast amount of data that is associated with a university and the variety of textual content. University websites, especially starting at the main site, consist of large html and text documents. This was perfect for this analysis as this research is only focused on text.

In configuring Nutch for crawling, no prefixes were excluded from the crawl. However, only html and text documents were pulled. A depth of 10 was used to limit the number of links the crawl would follow. A maximum number of 6,000 sites was used to limit the size of the data.

IV. PRELIMINARY METHODS FOR PULLING DEFAULTS

A. Scope

This research seeks to identify semantic defaults for verb events only. Any events that are not nouns are outside of the scope of this research and will be addressed in future work.

B. Procedures

1. Identify the most frequent 200 verbs from Brown corpus.

2. Pull Sentences using the Verb List. Once the top 200 verb stems are chosen, all of the sentences containing those verbs in any verb forms (for example, VBG, VBZ, VBP) are taken from Brown Corpus. This means that if a sentence had

 ISBN: 978-1-61208-390-2

the word *walking* tagged as a verb, the sentence was pulled. If the sentence had the word *walking* tagged as a noun, the sentence was not pulled. This resulted in 14719 sentences.

3. Create Dependency Representations. Using Stanford Parser, dependency representations were created for each sentence.

4. Select Noun and Adjective Arguments for Verb Events. For this initial analysis, we chose to select the following:

- all lone verbs;
- verbs with just nouns attached to them;
- verbs with adjectives and nouns attached.

As this was the only information needed from the dependency grammars, we chose to only select lines of the dependency grammars with the tags "nsubj", "dobj", "iobj" and "amod". The "nsubj tag was used to pull verbs, "dobj" and "iobj" were used in order to connect a verb to a noun ,or to pull verbs that did not have an "nsubj" tag, and "amod" was used to connect the verb-noun combinations to an adjective.

The dependency grammar below demonstrates how information was pulled:

nn(Brandt-2, Jackie-1)
nsubj(singled-3, Brandt-2)
nsubj(start-11, Brandt-2)
root(ROOT-0, singled-3)
advmod(singled-3, deep-4)
det(hole-7, the-6)
prep_into(singled-3, hole-7)
prep_at(hole-7, short-9)
aux(start-11, to-10)
xcomp(singled-3, start-11)
det(rally-13, the-12)
dobj(start-11, rally-13)

In this example, if *singled* were one of the top 200 verbs, it would be pulled because it has the "nsubj" tag. Originally we could classify it as a lone verb. *Start* would be pulled from the dependency grammar for the same reasons. "Start" and *rally* would also be pulled because of the "dobj" tag. However, because *start* is paired with *rally* in dobj, "start" would be removed from the lone verb list and added to the list of verbs with nouns.

5. Compare Verb-Noun List to Verb-Noun-Adjective List. In this step, we compared the list of verbs with nouns to the list of verbs with nouns and adjectives. As no default should appear unmodified, the verb-noun combinations from the verb-noun list were removed from the verb-noun-adjective list. Thus, if a noun occurred with a verb and had no modifier, it could not be considered a possible default for that event. For example, if we had the verb event *eat* we would see it with the unmodified noun *food food* is placed on the candidate list of defaults for *eat*. However, since we know that eating food is not informative, we don't expect to see the verb-noun pair (eat, food) to occur. It is entirely possible to say I eat hot food. This is because the heat of the food is relevant and not implied or assumed. Thus, we expect that it is possible to see the triple (eat, food, hot). If we saw the triple (*eat, food, hot*) in the corpus and never saw (*eat, food*) alone in the corpus, we would flag *food* as the default for *eat*. However, if we did see (*eat, food*) alone in the corpus then we would not consider *food* to be the default for *eat*.

V. Discussion of Preliminary Results

In pulling the relevant information from the dependencies, it was found that there were 13435 instances of lone verbs that map to events. The verbs with the highest frequency of lone occurrence included *made, come, felt, knew, began* and *look*. It was also found that there were 8240 verb-noun combinations and 2565 verb-noun-adjective combinations. Examples of verb-noun combinations included *reduce expense, create resources* and *wrote parts*. Examples of verb-adjective-noun combinations included *reported local romance, need new box* and feel questioning eyes. There were 2235 instances of candidate defaults found for 449 unique verb forms.

Although further analysis must be done, this seems consistent with [1-3] on defaults and the WD-Inference. According to their work, a verb-adjective-noun combination would likely represent the violation of a default. This is because the indication of additional detail in an event points to information that is out of the ordinary for the speaker or writer. As such, we would except to see the fewest instances of these combinations. We can see this in the example *feel questioning eyes*. The fact that the author needed to indicate that the eyes were questioning implies that something is out of the norm. This is consistent with how a native speaker would see that phrase. It is implied to the native speaker that *questioning eyes* are out of the ordinary.

The abundance of lone verbs is also consistent. When we say *I drove* the implication is that we drove a car. However, we don't actually mention the car unless it is out of the ordinary. For instance, if I am used to driving a motorcycle then it would be significant for me to say *I drove a car*.

VI. Next Steps

The next step in this research will be to apply the same methods used on Brown Corpus to the data that was pulled from the Purdue web crawl. The data will then need to be compared to the Brown sample to determine whether or not the findings from the structures are consistent.

In looking at the Web Crawl, we plan to pull data from not just the text documents themselves but also from image titles as well. It is possible that defaults and default violations will exist in this data.

Once both corpora have been examined, the verbs, nouns and adjectives will need to be acquired into the Ontology and Lexicon. Text Meaning Representations will then be generated for these sentences.

 ISBN: 978-1-61208-390-2

VII. FUTURE WORK

A. Implementation into OST

In order to fully implement default detection into OST, methods will need to be created for storing defaults. There will also need to be methods created for flagging defaults within a Text Meaning Representation. This may possibly require the creating of an InfoBase for each individual contributor.

B. Creation of InfoBase-like structures for individual defaults

As this is initial research concerning defaults, we are examing the defaults of a group of authors. Ideally, we need to be able to pull a set of defaults for a single author. As of now, we believe this will require individual InfoBases. InfoBases are meant to show the connections between several TMRs in order to create a larger picture of a conversation. Recording a series of defaults and default violations for an individual will help us better understand both what a person is saying and not saying in a conversation.

REFERENCES

[1] J. M. Taylor, V. Raskin, C. F. Hempelmann, and S. Attardo, "An unintentional inference and Ontological property defaults," Proc. IEEE_SMC, Istanbul, Turkey, 2010

[2] V. Raskin, J. M. Taylor, and C. F. Hempelmann, "Ontological semantic technology for detecting insider threat and social engineering," Pre-Proc. NSPW-10. Reprinted in: K. Beznosov, ed., Proceedings: New Security Paradigms Workshop 2010. September 20-23, 2019, Concord, MA, USA. New York: ACM Press, 2010

[3] V, Raskin, and J. M. Taylor, "A fresh look at semantic natural language information assurance and security: NL IAS from watermarking and downngrading to discovering unintended inferences and on to situational conceptial defaults," in: B. Akhgar and H. R. Arabnia, eds., Emerging Trends in Informationa nad Communication Technologies Security. Amsterdam: Elsebvier (Morgan Kaufmann, 2013

[4] V. Raskin, C. F. Hempelman, and J. M. Taylor, "Guessing vs. knowing: The two approaches to semantics in natural language processing,: Proc. Dialogue 2010. Bekasovo/Moscow, Russia, pp. 642-650, 2010

[5] J. M. Taylor, C. F. Hempelmann, and V. Raskin, "On an automatic acquisition toolbox for ontologies and lexicons in ontological semantics," Proc. ICAI-10, Las Vegas, NE, pp. 863-869, 2010

[6] J. M. Taylor, and V. Raskin, Understanding the unknown: Unattested inpiut in natural language," Proc. FUZZ-IEEE-11. Taipei, Taiwan 2011

[7] C. F. Hempelmann, J. M. Taylor, and V. Raskin, "Application-guided ontological engineering, " Proc. ICAI-10. Las Vegas, NE, 2010

[8] J. M. Taylor, V. Raskin, and C. F. Hempelmann, "From disambiguation failures to common-sense knowledge acquisition: A day in the life of an ontological semantic system," Proc. WI-IAT-11. Lyon, France, 2011

[9] J. M. Taylor, V. Raskin, and C. F. Hempelmann, "Towards computational guessing of unknown word-meanings: The ontological semantic approach," Proc. CogSci-11. Boston, MA 2011

[10] S. Nirenburg, and V. Raskin, Ontological Semantics. Dordrecht: D. Reidel, 2004.

[11] H. P. Grice, "Logic and conversation," in: P. Cole and J. L. Morgan, eds., Syntax and Semantics, Vol. 3, Speech Acts. New York: Academic Press, 1975

[12] M. J. McShane, A Theory of Ellipsis. New York: Oxford University Press, 2005

[13] J. Manvika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, C., and A. H. Byers, Big Data: The Next Frontier for Innovation, Competition, and Productivity.

[14] Big data. *IT Glossary*, in Gartner: Retrieved November 21, 2014, from http://www.gartner.com/it-glossary/big-data, 2014

[15] W.N. Francis, and H. Kucera, (1979). Brown corpus manual. Brown University.

[16] M.C. De Marneffe, and C.D. Manning, (2008). Stanford typed dependencies manual. URL http://nlp. stanford. edu/software/dependencies manual. pdf.

 ISBN: 978-1-61208-390-2

Generating Opinion Agent-based Models by Structural Optimisation

A.V. Husselmann

Computer Science, Massey University
Email: a.v.husselmann@massey.ac.nz

***Abstract*—Agent-based modelling has enjoyed a significant increase in research effort in recent years. Particular efforts in the combination of it with optimisation algorithms have allowed the automatic generation of interesting system-level behaviors. The vast majority of these efforts have focussed on parametric optimisation, whereby the structure of a model remains user-defined, and parameters are systematically calibrated. Fairly little effort has been expended in investigations towards combinatorial optimisation in the context of agent-based modelling. The author has previously shown that it is possible to combine the use of a domain-specific language (DSL) and the multi-stage programming paradigm to provide a platform suitable for extension using an evolutionary algorithm. This combination was important to allow run-time code generation, while using just-in-time compiling. The entire process is extremely compute intensive, but can be successfully mitigated using such a language. In this article, three experiments are carried out using this language, and the efficacy of this optimisation is discussed.**

Keywords–Agent-based models; Optimisation; Domain-specific languages; Multi-stage programming.

I. INTRODUCTION

Agent-based modelling (ABM) is significantly multi-disciplinary. It has been used to describe many models which are in essence intuitively reduced to local interactive behavior, which compound over time to generate macro-level phenomena, which are not necessarily specified [1], [2]. Such models famously include Reynolds' "Boids" [3], in which simple interactive rules generate complex behavior reminiscent of flocking birds and schooling fish. Some disciplines which have successfully made use of ABM include medicine [4], [5], political science [6], microbiology [7], and social science more generally [8], [9], as well as extensively across ecology [10], [11], [12].

Models of opinion [13], [14] are particularly well-suited to being studied using ABM. Simple implementations of these models are considered briefly in this article for aiding in optimisation. These include a very simple voter model [15], the Sznajd opinion model [14], and Axelrod's model of cultural dissemination [6]. These familiar models were used in this work as a basis for automatically obtaining new opinion-based models for accomplishing certain objectives given by scalar objective functions.

Domain-specific Languages (DSLs) are making the use of ABM much more streamlined. While no formal definition exists, DSLs are essentially compiled or interpreted languages made specifically for a small target application domain [16]. It is generally agreed that a DSL should be well defined in terms of target domain, syntax, as well as formal, and informal semantics [17]. Indeed, DSLs have already made way into the field of ABM, such as the recent work of Franchi, who presented a DSL built on Python for agent-based social network modelling [18], as well as the NetLogo modelling environment [19]. Typically, ABM modelling packages rely on more general purpose languages such as Java [20], [21].

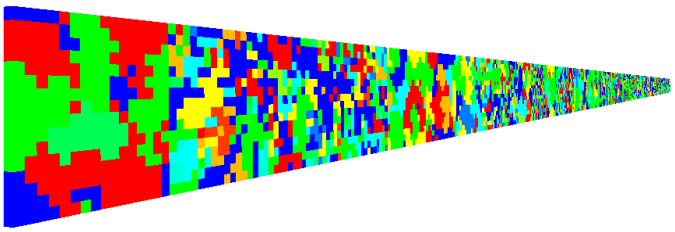


Figure 1. 64, 16x16 side-by-side recombined opinion-based lattice models being evaluated.

An excellent method of creating DSLs while keeping them extensible and fast [22], is the Multi-stage Programming (MSP) paradigm [23]. MSP is particularly attractive for its ability to avoid penalties for run-time code generation (RTCG) [22]. An MSP-based language released recently in 2013 is Terra [24], invented by DeVito et al. It was positioned as the lower-level, high performance counterpart to Lua, a loosely-typed general purpose scripting language [25], [26]. Terra makes use of LuaJIT, which is a just-in-time (JIT) compiler for the Lua language [27]. Terra's implementation of MSP allows the user to write a full compiler architecture in Lua to parse a DSL and generate programs by splicing together Terra code fragments, which are ultimately JIT-compiled before being executed at will.

A DSL (code named MOL) was created using Terra, for the purpose of agent-based model induction, and was presented mid-2014 [28]. This language combines techniques from Gene Expression Programming with the multi-stage paradigm of Terra. It accomplishes this using several stages, involving a parser, type checker, DSL-optimizer, and code generator all written in Lua (and heavily inspired by examples distributed with Terra [24]), with a run-time generator and finally a main Lua program to execute the runtime.

Novelty arises from the combination of this language with a combinatorial optimizer. Genetic Algorithms [29] operate by evaluating a fitness function for a set of candidates in a population [30], and subsequently the population is modified by genetic operators to converge upon an optimal candidate. Similarly here, a population of candidates is evaluated using the runtime, and the main Lua program is then tasked with using a Lua-based optimizer to manipulate candidate programs and recompile them using the DSL compiler stack, and runtime generator. The result of the simulation when compiled with a user interface, is shown in Figure 1. This figure shows multiple opinion-based models being evaluated, where each

 ISBN: 978-1-61208-390-2

model differs from the next in subtle or extreme ways. The best model would be found, and used to generate new models which better suit the objective function provided by the user. The language itself is capable of generating models which are structurally different: in contrast to optimizers which generate models with different parameters.

In the next section, related research and rationale for the approach outlined in this research is presented. In Section III, more detail on the DSL is provided, including its optimisation algorithm. A methodology is then given for evaluating the ability of the language to search through the space of agent-based models for the purpose of inducing some new models. Results of these are given in Section IV. A discussion is provided in Section V. Finally, the article is concluded in Section VI with some areas for future work.

II. Related Research

Previous research attempts to optimise the *structure* of agent-based models using optimisation have not involved DSLs. The recent work of Van Berkel [31], [32] in 2012 involved the use of Grammatical Evolution [33], in order to generate NetLogo [19] programs from predetermined building blocks. While a sophisticated approach, it did not involve a dedicated DSL. A DSL would allow one to prototype different approaches more quickly, rather than re-engineering an existing code base for a different model. Moreover, run-time interpretation of candidate solutions present a significant performance overhead.

Junges and Klügl in 2010, investigated the problem using learning classifier systems, Q-learning and Neural Networks for generating behavior [34]. This was followed in 2011 by their investigation with Genetic Programming [35]. No clear winner among these algorithms was drawn out by the authors, however, they did note in 2012 that Reinforcement Learning and Genetic Programming proved to be more suitable, as they generate human-readable results [36]. It seems appropriate in these circumstances to allow the end-user of such an optimizer a larger breadth of control over the algorithms, whilst still ensuring that it is simple enough to use. This is precisely what Multi-stage Programming allows one to accomplish.

Earlier in 2002, Privošnik developed an evolutionary optimizer which evolved agents with customised finite state machines to solve the Ant Hill problem [37]. While sophisticated, no indication was given concerning reuse of the system for other models. The work of Junges and Klügl however, was integrated with the SeSAm platform to provide a method for use by other researchers [35]. While this is certainly encouraging, the problem of severe performance inadequacies is frequently overlooked. None of these works (except for Van Berkel [31]) extensively consider performance difficulties.

Performance is a very significant issue, and if left unmitigated, can undermine even a sophisticated optimizer. The approach taken in the method described in the next section attempts to solve two problems. One of mitigating excessive computation using parallelism, and the other of still allowing simple interaction with a sophisticated optimizer. The method described is the only known agent-based modelling language with an embedded optimizer, which compiles without alteration to both graphics processing units (GPUs) and single-threaded code without modification. In addition, thanks to Terra and LLVM ("low level virtual machine" compiler architecture) this approach does not suffer run-time interpretation costs, due to generated code being compiled and executed at run-time, for both GPU and CPU. The purpose of this work is to demonstrate and evaluate how well this approach can generate opinion models given a specific objective.

III. Method

The DSL compiler architecture introduced in Section I involves several stages. A flow diagram is provided in Figure 2 which details the process in which special DSL code is compiled. Part of this flow diagram is repeated during run-time, allowing for run-time code generation (RTCG). The process involves three distinct stages. The first is a custom compiler architecture written in Lua, which compiles DSL code to Terra code. Then, using multi-stage programming operators, the Terra code is spliced into a simulation program, and finally compiled using the usual internals of Terra, which involves LLVM. The lattice on which the program operates is updated by this final program exactly once, depending on what update scheme is selected. For example, one may choose between a Monte Carlo style update, or a red-black update style. Clearly these involve different code structures, but fortunately, it is notably easy to accomplish this using the MSP paradigm.

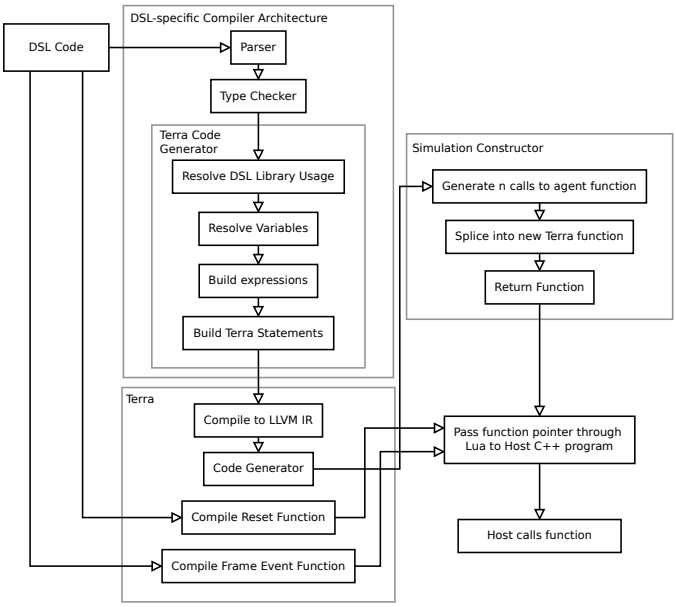


Figure 2. A flow diagram indicating compile-time and some run-time flow.

The final compiled runtime is embedded in a Lua script, which executes the compiled code up to t_n timesteps at a time, before (optionally) rendering the resulting simulations, and at certain points, regenerate all programs by passing the programs through the optimizer, and through the rest of the compiler stack before re-executing the set of programs to obtain new fitness values. The optimizer simply modifies specially marked parts of the Terra expression trees, which are stored in a Lua datastructure.

It has been previously shown that this language and architecture is capable of compiling to NVIDIA PTX code and fully exploit graphics card processing power [28], [38], [39]. The purpose of this article is to cast further light on the

 ISBN: 978-1-61208-390-2

optimisation characteristics of the language, as opposed to its ability to compile to different architectures. Here, the only runtime used is the CPU-based runtime. The complete process is summarised in the algorithm in Figure 3.

```
Allocate large lattice on host
Divide lattice into n rectilinear segments

Parser constructs typed tree from agent prototype code
Type checker checks programs
Platform code generator creates a Terra function f*
Optimiser duplicates f* n times
Simulator code generator creates a stepping function:
for i=0,n do
    Simulator generator inserts call to function i into Terra
    statement list L
end for
L is the simulator Terra code fragment

Runtime generator creates F_r, the runtime Terra function
Inserts L into F_r
Compile F_r (compiles all n Terra trees to machine code)

Begin Lua main program:
Reset fitness scores, reset lattices

for g=0,generation_count do do
    Create Terra program function F_P
    F_P: Insert fragment to zero scores for models
    F_P: Add for-loop for_repeat to F_P for every repeat
    F_P: Insert fragment to reset lattices in for_repeat
    F_P: Insert render call in for_repeat
    F_P: Insert fragment to swap lattices
    F_P: Add for-loop for_timestep to for_repeat
    F_P: Insert call to F_r in for_timestep
    F_P: Insert fragment to swap lattices
    F_P: Insert render call in for_timestep
    Compile the Terra function F_P
    Execute F_P
    Pass programs through optimiser with scores
    Optimiser: Selection (program constructs)
    Optimiser: Recombine selected constructs
    Optimiser: Mutate constructs
    Pass modified programs to simulator code generator
    Pass simulator code to runtime generator
    Recompile the Terra function F_r
    Reset lattices, scores
end for
```

Figure 3. An algorithm describing a complete run of the system for an optimisation test.

As shown in Figure 3, all code generated using the system is generated as fragments of Terra, and spliced together in various stages and various configurations. Once a monolithic Terra function is fully generated, the Terra compiler is used to compile the code at runtime to machine code, using LLVM [40]. When run with the user interface enabled, the compiled program is capable of requesting a re-draw of the simulators. The function is emptied when the user interface is disabled.

The optimisation phase involves a simplified version of the Gene Expression Programming (GEP) algorithm of Ferreira [41], [42]. Whereas the GEP algorithm is designed with several operators for circulating information throughout candidate programs, the recombination optimizer in this system is restricted to a very simplistic minimal set of genetic operators: mutation, crossover and selection. This is purely for convenience at this stage, and will be subject of considerable future improvement.

GEP and the simplified algorithm used here both involve the use of candidates represented as strings of symbols or "codons". It was necessary to be able to translate from a program abstract syntax tree (AST) to a string of codons. The method used to accomplish this is the Karva language, also designed by Ferreira, for the use of GEP [42]. For brevity, this language is not discussed here in great detail. The reader is kindly referred to the excellent book on GEP by Ferreira [42]. Karva expressions, or *k*-expressions, are composed of a head section and tail section. The head-section may contain any symbols, and the tail section may only contain terminal symbols. Terminal symbols are effectively statements which do not involve control flow. Non-terminal symbols correspond to fragments of code which involve control blocks such as *if*-statements. Two additional non-terminals were provided: *P0* and *P1*. These simply execute their arguments sequentially. *P0* is of arity 2, and *P1* is of arity 3.

Three experiments were carried out to cast light on the efficacy of the modelling approach introduced here. Each experiment involved a population of 64 simulations, each one operated on a regular lattice of size 16 by 16. A single evaluation of a model with respect to the objective functions provided was obtained after 50 timesteps, and averaged over 20 (unless otherwise noted) independent runs. The maximum number of evaluation runs (or "generations") was set at 100 during all tests. This number was found to be empirically adequate for obtaining meaningful results, though it would be useful to know in future work how this algorithm behaves with longer runs.

Furthermore, the head length of the *k*-expressions was 3. The update style of the lattices was two-phase with a randomised order, ensuring that all lattice sites executed at least once during a single timestep.

```
select recombination to maximise(score)
 -- pick a new neighbour
 select all
  idx = get_random_neighbour
  s = getneighbour(lattice,idx)
 end

 -- Snzajd Model
 if (me - s) == 0 then
  propagate_n6(newlattice,idx,posx,posy,posz,me)
 end

 -- Axelrod Model
 if randomfloat > (abs(s-me) / OPINIONCOUNT) then
  propagate_single(newlattice,idx,me)
 end

 -- Voter Model
 propagate_single(newlattice,idx,me)
end
```

Figure 4. An excerpt from within the MOL program used in experiments.

 ISBN: 978-1-61208-390-2

Figure 4 contains a fragment of MOL DSL code from within the model code. For brevity, the rest of the model is omitted. Several functions were implemented specially for this model, using an extension framework. Function calls in this code are resolved to functions written in Terra itself in the same scope of the code. get_random_neighbour, however, is a macro, written in Lua, which generates Terra code while having direct access to environment variable references as obtained by the parser.

IV. Experiments

The experiments discussed in the next three sections involve the same code shown in Figure 4, except for the first line. The maximise keyword is used to indicate to the MOL compiler that a score is to be maximised by the optimizer. The expression in brackets, score, is essentially the objective function. The score can be modified at any timestep by any cell agent's program using the special environment variable timestep. The MOL compiler inserts a special statement in the program, which will add the computed score to the current simulation's score. This accumulated score is then later passed to the optimizer for processing.

Each experiment differs in objective function. The first experiment attempts to recombine the Axelrod, Sznajd, and Voter models in order to find a model which minimises the prevailing opinions in the models. That is to say, to find a model with homogeneous agents which converges to a single opinion in 50 time steps. The maximum time steps for this test was set to 50. The second experiment attempts to maximise a cumulative score $s = \sum_{i=0}^{n}(o_i)^{-1}$ where o_i is the number of opinions at timestep i. The third and final model attempts to maximise a more complicated objective function, which involves computing the standard deviation of scores, where each score represents a value indicating the distance from a fully opinion-equalised model.

A. Experiment 1 – Fast Consensus

In this experiment, the objective was to *minimise* the opinion count (ie. obtain consensus) within 50 timesteps. Unlike experiments 2 and 3, this experiment *minimises* the objective function. Here, the objective function is computed as $s(t_n) = o_i$ where o_i is the number of opinions at timestep i, and t_n is the final timestep of the evaluation run. While the compiler will be accumulating all values of $s(t)$ from t_0 to t_n, only $s(t_n)$ will be non-zero for this test. This automatic accumulation is quite useful for quick fitness objectives.

The fitness plot of this experiment is shown in Figure 5. Mean and minimum data are shown. Error bars on the mean indicate the standard deviation of the fitness values in the population. Interesting in this model is the fact that the first generation of programs actually contained a program capable of converging to roughly 4 opinions in 50 timesteps. After approximately 65 generations, a model was generated which could converge to two opinions in 50 timesteps. The results for the run are statistically significant due to the averaging occurring within every evaluation stage. However, it should be noted that this is one sample run of the entire system, and other runs may differ slightly. The purpose in providing sample runs is to qualitatively compare different model structure optimisation approaches particularly by means of different objective functions.

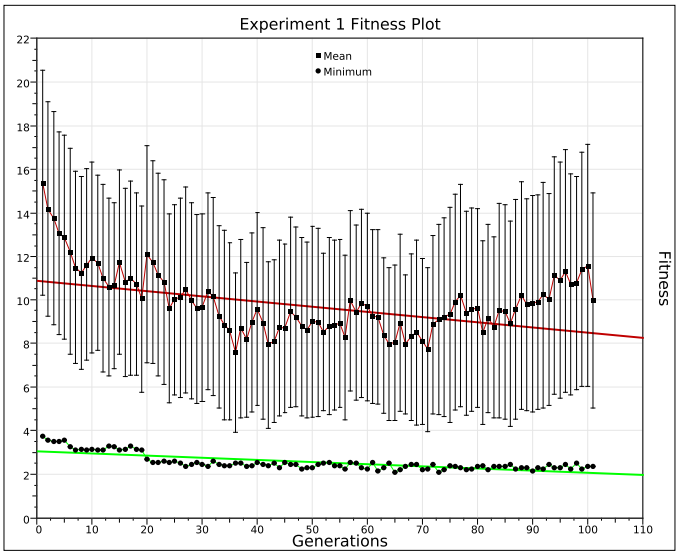


Figure 5. Fitness results by generation for experiment 1.

As seen in Figure 5, search stagnation prevents the algorithm from generating a model which converges to a single opinion. This could also be due to the specification of the model code in the optimisation structure. This is a significant disadvantage of the MOL language. It is possible that a user may provide insufficient information in the structure, and therefore the optimizer can never reach an optimal result. It is also possible that 50 timesteps is simply insufficient for reaching consensus in the system.

B. Experiment 2 – Cumulative Fitness

Experiment 2 involves the maximisation of a cumulative objective function, where the score is computed by successively adding the inverse of the number of opinions. Here, the optimizer favors models which eliminate opinions as quickly as possible; essentially related to experiment 1, except that it is *maximising* a *cumulative* score.

Figure 6 presents the results for this sample run. As before, each generation is averaged over 20 separate executions. The optimizer is able to maintain a good diversity in population, as is demanded of evolutionary algorithms such as these. Even though very simplistic genetic operators were used, an increase in fitness is observed from maximum fitness values. While maximum fitness increases (albeit slowly), the mean fitness remains approximately the same.

In similar fashion to experiment 1, Figure 6 indicates a relatively quick improvement in fitness, which is met after approximately 10 time steps with what appears to be search stagnation. It is possible that the optimizer has simply reached a set of candidates with higest fitness possible and is oscillating between them (causing the variation towards the latter 80 generations). The first 20 frames is, however, convincing as to the optimizer's ability.

C. Experiment 3 – Turbulence

The third experiment attempts to induce some turbulence by favoring models which have a high standard deviation for a fitness value comprised of a distance to equilibrium among

 ISBN: 978-1-61208-390-2

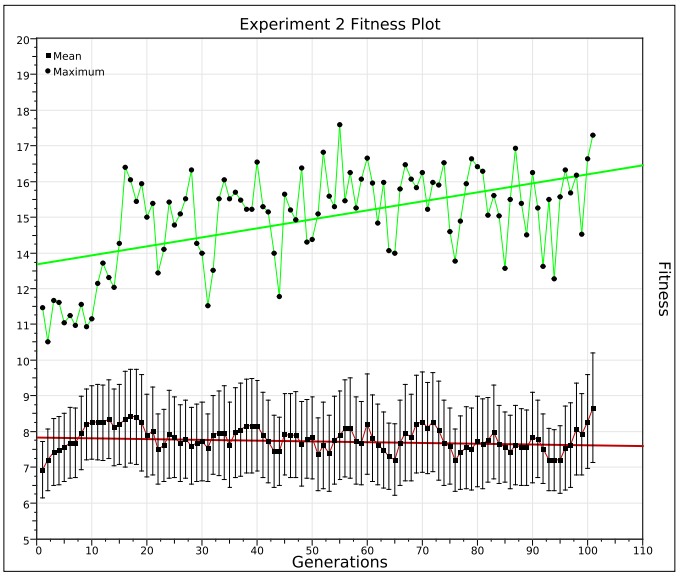


Figure 6. Fitness results by generation for experiment 2.

opinion counts. This equilibrium distance is computed by first histogramming the number of opinions in the lattice, and then summing the deviation of each opinion count from the number expected in order to form an equilibrium where every opinion has an equal share of lattice sites. This value is saved for every time step, and a standard deviation is computed on the very last time step, and added to the fitness score of the model in question. The objective function is therefore the standard deviation of the sequence:

$$s_i = o_c(i,j) - (o_n/16^2) \quad (1)$$

where $i = 1..n$, number of simulations is $n = 64$, $j = 1..o_n$, the number of opinions is $o_n = 32$, and $o_c(i,j)$ is the number of agents with opinion j in simulation (or candidate) i.

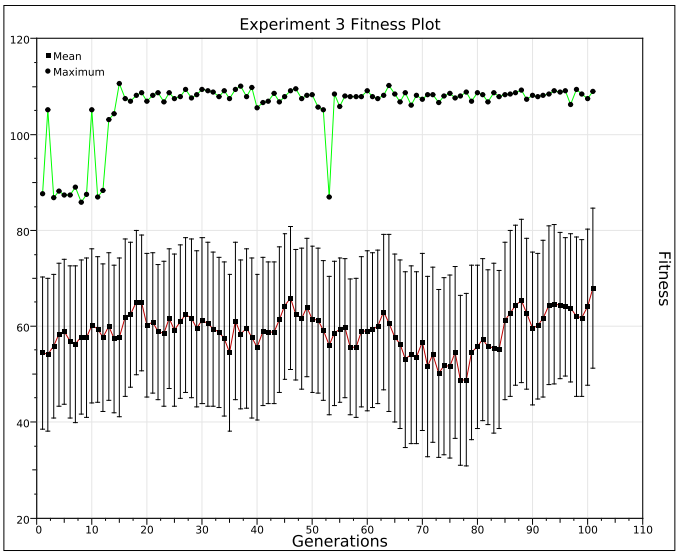


Figure 7. Fitness results by generation for experiment 3.

Figure 7 shows the result of this experiment as a fitness plot by generation. The purpose of the experiment is to determine the efficacy of the system as a whole, when dealing with significantly more complex objective functions. Shown in this figure is a sudden increase in fitness at approximately generation 15, which is reminiscent of the problem of local minima in metaheuristics [43].

To illustrate the effectiveness of the system, the best individual generated by the optimizer will be examined. The best program generated had a fitness score of 107, and had the optimisation construct composed of the following *k*-expression:

```
0 1 2 3 4 5 6 7 8 9 0
I0P0N1L0N1L0L0N0N0N2N2
```

For the purpose of the optimizer, the symbols are simply defined by assigning a type, and incrementing a number. Therefore, the first `if`-statement in Figure 4 is translated into `I0`, the second `if`-statement is translated into `I1` and so on, similarly for different statements. The terminal statements are identified by symbols which begin with `N` and `L`. `P0` is a manually inserted nonterminal, which simply executes its two arguments sequentially.

```
if (me − s) == 0 then
 select all
  idx = get_random_neighbour()
  s = getneighbour(lattice,idx)
 end
 propagate_single(newlattice,idx,me)
else
 propagate_single(newlattice,idx,me)
end
```

Figure 8. The best candidate generated to optimise the objective function provided for experiment 3. The *k*-expression for this code is I0P0N1L0N1L0L0N0N0N2N2.

When interpreted, the *k*-expression translates into the code shown in Figure 8. By inspection, this code does indeed provide turbulence. It is composed of a single if statement, which will replace the model code fragment in Figure 4 as part of the model code. Should the condition be true, then the randomly chosen neighbor shares the same opinion. If this is the case, then firstly, the simulation will choose a *different* neighbor, and propagate their opinion to that neighbor. Note that in the second case, the simulation does not check the opinion of the neighbor chosen. Should the condition be false, it signifies that the neighbor chosen does not hold the same opinion, and it would therefore propagate its opinion to that neighbor.

Since the lattice update style relies on a randomised-order update sequence between lattice timesteps, each cell is guaranteed to execute their program in a timestep. However, the final result does indeed depend on what order agents in the lattice are updated. Therefore, the result of the optimisation run may also depend on the update style. If it were a *monte-carlo* style update, then not all cells may be executed, and some may be executed twice.

The result shown in Figure 4 does indeed provide turbulence in the model by attempting to guarantee that an agent will propagate its opinion. It may be possible there is a program which provides more turbulence, however the optimizer in

 ISBN: 978-1-61208-390-2

this case was limited to three non-terminal statements in a candidate. In this example, only two were used: an `if`-statement, and a `P0` statement. More complex solutions may improve upon this fitness value, but do potentially require more computation.

V. Discussion

The three experiments conducted show that model programs can be optimised for different purposes. The model strategy that was being investigated here assumed that the modeller was attempting to use an optimizer to generate micro-behaviors that they are interested in. The results appear to show that it is indeed useful to obtain results which could be useful in a modelling situation. Exactly how useful the results are, would depend on the quality of the objective function, and terminals and non-terminals, which are discussed below.

There are unfortunately some caveats which are associated with Genetic Programming and general evolutionary algorithm literature [44], [45]. Particularly, the choice of terminals and nonterminals is a problem that is still applicable in the system discussed in this article. The implications of this mean that what is considered "optimal" by the optimizer, may in fact be severely limited by a wrong choice in initial program. The code shown in Figure 4 is very important, not in order and exact syntax, but more in terms of abstract states, lattice modifications, and state transitions. It is for this reason that future work will likely involve the design of a second DSL, which will handle finite state machines separately.

The optimizer also depends on the user for appropriate selection of parameters. Like many optimisation algorithms, a set of parameters is necessary. In the case of this system, probabilities of mutation, crossover and selection are predefined and hand calibrated. Clearly, some parameters would suit better in different situations. By considering extreme values in these parameters, it is easy to see how the algorithm would fail: setting the mutation probability to zero will remove all chances of injecting new material into the population of candidates, and therefore only "genetic drift" will occur [42], and similar problems will occur with the other parameters. Therefore, at least sensible generic values are absolutely necessary.

In particular, a parameter which is of particular importance here, pertains to the *k*-expressions of the candidate program population. One must choose a suitable head length for candidates, and then an expression length can be inferred from this to ensure that all generated programs are valid. This problem is loosely associated with the problem of avoiding code bloat in genetic algorithms [44].

At this point then, with the limitations of this approach, it may only be suitable to a small subset of models and their development. This is made clearer by the demand for an optimisation function, provided by the user. Such a function is not easy to formulate, and in many cases, generates results that appear to exploit a subtle flaw.

VI. Conclusion and Future Work

This article has introduced a proof of concept language and optimizer system presented in mid-2014, and provided some experimental results to indicate its efficacy. A model was written in this language, and a portion of "uncertain" code was also written with simplistic implementations of the Sznajd, Axelrod and Voter models. Three different optimisation functions were used, in order to instruct an optimizer on how to recombine the code given. The novelty in this approach lies in its use of a multi-stage language which is capable of run-time code generation using LLVM, and thereby avoiding costly run-time interpretations of code as many evolutionary algorithms do. Sample runs were made, and results presented in the form of fitness by generation plots.

The first experiment's optimisation function intended to minimise the number of opinions in the model. While each model was 16 cells by 16 cells, there were up to 32 opinions and 64 candidate models in the population. A limited improvement in minimum fitness was achieved. The second experiment attempted maximisation of a cumulative fitness function. It was intended that maximising this quantity would produce a model that attempts to reach consensus (minimum number of opinions) as quickly as possible. For brevity, a thorough examination of the best individual was omitted.

The third experiment involved a more complex fitness function, to determine how the system handles nontrivial objective functions. In short, an improvement in fitness was observed and the examination of the best individual provided some insights into the optimizer's behavior.

To conclude, the experiments appear to show a process that would be useful in a modelling situation, albeit, a tradeoff is presented in which the user must be able to provide a candidate solution, as well as an objective function. The quality of these directly influence the quality of the results, and if not adequately specified, may not be useful at all. These problems appear to stem from fundamental issues in the Genetic Programming and Evolutionary Algorithm literature.

Future work on this system involves a thorough statistical study by running the simulation on Graphical Processing Units (GPUs), and providing robust indications of both large-scale model optimisation, as well as small-scale mean performance. Further improvements on the optimizer itself is also under consideration.

References

[1] E. Bonabeau, "Agent-based modeling: Methods and techniques for simulating human systems," Proceedings of the National Academy of Sciences of the United States of America, vol. 99, no. Suppl 3, 2002, pp. 7280–7287.

[2] C. M. Macal and M. J. North, "Tutorial on agent-based modeling and simulation part 2: How to model with agents," in Proc. 2006 Winter Simulation Conference, Monterey, CA, USA, 3-6 December 2006, pp. 73–83, iSBN 1-4244-0501-7/06.

[3] C. Reynolds, "Flocks, herds and schools: A distributed behavioral model," in SIGRAPH '87: Proc. 14th Annual Conf. on Computer Graphics and Interactive Techniques, Maureen C. Stone, Ed. ACM, 1987, pp. 25–34, iSBN 0-89791-227-6.

[4] G. P. Figueredo, P.-O. Siebers, and U. Aickelin, "Investigating mathematical models of immuno-interactions with early-stage cancer under an agent-based modelling perspective," BMC Bioinformatics, vol. 14, no. 6, 2013, pp. 1–38.

[5] G. P. Figueredo, P.-O. Siebers, U. Aickelin, and S. Foan, "A beginner's guide to systems simulation in immunology," in Artificial Immune Systems, ser. Lecture Notes in Computer Science, C. A. Coello Coello, J. Greensmith, N. Krasnogor, P. Liò, G. Nicosia, and M. Pavone, Eds. Springer Berlin Heidelberg, 2012, vol. 7597, pp. 57–71.

[6] R. Axelrod, "The dissemination of culture: a model with local convergence and global polarization," J. Conflict Resolution, vol. 41, 1997, pp. 203–226.

 ISBN: 978-1-61208-390-2

[7] F. L. Hellweger, E. S. Kravchuk, V. Novotny, and M. I. Glayshev, "Agent-based modeling of the complex life cycle of a cyanobacterium (anabaena) in a shallow reservoir," Limnology and Oceanography, vol. 53, 2008, pp. 1227–1241.

[8] J. M. Epstein and R. Axtell, Growing Artificial Societies : Social Science From the Bottom up., ser. Complex Adaptive Systems. Brookings Institution Press, 1996.

[9] J. M. Epstein, "Agent-based computational models and generative social science," Generative Social Science: Studies in Agent-Based Computational Modeling, 1999, pp. 4–46.

[10] J. Ferrer, C. Prats, and D. López, "Individual-based modelling: An essential tool for microbiology," J Biol Phys, vol. 34, 2008, pp. 19–37.

[11] V. Grimm and S. F. Railsback, Individual-based Modeling and Ecology. Princeton University Press, 2005.

[12] D. Helbing and S. Balietti, "How to do agent-based simulations in the future: From modeling social mechanisms to emergent phenomena and interactive systems design," Santa Fe Institute, Tech. Rep., 2011.

[13] K. Kacperski et al., "Opinion formation model with strong leader and external impact: a mean field approach," Physica A: Statistical Mechanics and its Applications, vol. 269, no. 2, 1999, pp. 511–526.

[14] K. Sznajd-Weron and J. Sznajd, "Opinion evolution in closed community," International Journal of Modern Physics C, vol. 11, no. 06, 2000, pp. 1157–1165.

[15] T. M. Liggett, Stochastic interacting systems: contact, voter and exclusion processes. Springer, 1999, vol. 324.

[16] D. Spinellis, "Notable design patterns for domain-specific languages," Journal of Systems and Software, vol. 56, 2008, pp. 91–99.

[17] W. Taha, "Domain-specific languages," in Pro. Int. Conf. Computer Engineering and Systems (ICCES), 25-27 November 2008, pp. xxiii – xxviii.

[18] E. Franchi, "A domain specific language approach for agent-based social network modeling," in IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2012, pp. 607–612.

[19] S. Tisue and U. Wilensky, "NetLogo: A simple environment for modeling complexity," in International Conference on Complex Systems, 2004, pp. 16–21.

[20] N. Collier, "Repast: An extensible framework for agent simulation," Social Science Research Computing, University of Chicago, Tech. Rep., 2003.

[21] S. Luke, C. Cioffi-Revilla, L. Panait, K. Sullivan, and G. Balan, "MASON: A multiagent simulation environment," Simulation, vol. 81, 2005, pp. 517–527.

[22] W. Taha, "A gentle introduction to multi-stage programming," in Domain-Specific Program Generation. Springer, 2004, pp. 30–50.

[23] W. Taha and T. Sheard, "Multi-stage programming with explicit annotations," in ACM SIGPLAN Notices, vol. 32, no. 12. ACM, 1997, pp. 203–217.

[24] Z. DeVito, J. Hegarty, A. Aiken, P. Hanrahan, and J. Vitek, "Terra: a multi-stage language for high-performance computing." in PLDI, 2013, pp. 105–116.

[25] R. Ierusalimschy, L. H. de Figueiredo, and W. Celes, "The evolution of Lua," in Proceedings of the third ACM SIGPLAN conference on History of programming languages. ACM, 2007, pp. 2–1–2–26.

[26] R. Ierusalimschy, L. H. de Figueiredo, and W. C. Filho, "Lua: An extensible extension language," Software: Practice and Experience, vol. 26, 1996, pp. 635–652.

[27] M. Pall, "The luajit project," 2008. [Online]. Available: www.luajit.org

[28] A. V. Husselmann, "Data-parallel structural optimisation in agent-based modelling," Ph.D. dissertation, Massey University, 2014.

[29] J. H. Holland, Adaptation in natural and artificial systems. Ann Arbor: University of Michigan Press, 1975.

[30] D. E. Goldberg, Genetic Algorithms in Search, Optimization & Machine Learning. Addison-Wesley Publishing Company Inc, 1989.

[31] S. van Berkel, "Automatic discovery of distributed algorithms for large-scale systems," Master's thesis, Delft University of Technology, 2012.

[32] S. van Berkel, D. Turi, A. Pruteanu, and S. Dulman, "Automatic discovery of algorithms for multi-agent systems," in Proceedings of the fourteenth international conference on Genetic and evolutionary computation conference companion, July 2012, pp. 337–334.

[33] C. Ryan, J. Collins, and M. O'Neill, "Grammatical evolution: Evolving programs for an arbitrary language," in Proceedings of the First European Workshop on Genetic Programming, ser. LNCS, vol. 1391. Paris: Springer-Verlag, April 1998, pp. 83–95.

[34] R. Junges and F. Klügl, "Evaluation of techniques for a learning-driven modeling methodology in multiagent simulation," in Multiagent System Technologies, ser. Lecture Notes in Computer Science, J. Dix and C. Witteveen, Eds. Springer Berlin Heidelberg, 2010, vol. 6251, pp. 185–196.

[35] R. Junges and F. Klügl, "Evolution for modeling - a genetic programming framework for SeSAm," in 13th Annual Genetic and Evolutionary Computation Conference, GECCO 2011, Proceedings, 2011.

[36] ——, "Programming agent behavior by learning in simulation models," Applied Artificial Intelligence, vol. 26, 2012, pp. 349–375.

[37] M. Privošnik, M. Marolt, A. Kavčič, and S. Divjak, "Construction of cooperative behavior in multi-agent systems," in Proceedings of the 2nd International Conference on Simulation, Modeling and optimization (ICOSMO 2002). Skiathos, Greece: World Scientific and Engineering Academy and Society, 2002, pp. 1451–1453.

[38] A. V. Husselmann and K. A. Hawick, "Towards high performance multi-stage programming for generative agent-based modelling," in INMS Postgraduate Conference, Massey University, October 2013.

[39] ——, "Multi-stage, high performance, self-optimising domain-specific language for spatial agent-based models," in The 13th IASTED International Conference on Artificial Intelligence and Applications. Innsbruck, Austria: IASTED, February 2014.

[40] C. Lattner and V. Adve, "Llvm: A compilation framework for lifelong program analysis & transformation," in Code Generation and Optimization, 2004. CGO 2004. International Symposium on. IEEE, 2004, pp. 75–86.

[41] C. Ferreira, "Gene expression programming: A new adaptive algorithm for solving problems," Complex Systems, vol. 13, no. 2, 2001, pp. 87–129.

[42] ——, Gene Expression Programming, 2nd ed., ser. Studies in Computational Intelligence, P. J. Kacprzyk, Ed. Berlin Heidelberg: Springer-Verlag, 2006, vol. 21, iSBN 3-540-32796-7.

[43] A. V. Husselmann and K. A. Hawick, "Levy flights for particle swarm optimisation algorithms on graphical processing units," Parallel and Cloud Computing, vol. 2, no. 2, April 2013, pp. 32–40.

[44] R. Poli, L. Vanneschi, W. B. Langdon, and N. F. McPhee, "Theoretical results in genetic programming: the next ten years?" Genetic Programming and Evolvable Machines, vol. 11, 2010, pp. 285–320.

[45] T. Weise, M. Zapf, R. Chiong, and A. J. Nebro, "Why is optimization difficult?" in Nature-Inspired Algorithms for Optimisation, SCI 193. Springer-Verlag, 2009, pp. 1–50.

 ISBN: 978-1-61208-390-2

Modeling and Studying Cooperative Behavior between Intelligent Virtual Agents by Means of PRE-ThINK Architecture

Dilyana Budakova
Department Of Computer Systems and Technologies
Technical University of Sofia, Branch Plovdiv, Plovdiv, Bulgaria
dilyana_budakova@tu-plovdiv.bg

Lyudmil Dakovski
Department of Communication & Computer Technologies
European Polytechnic University
Pernik, Bulgaria
lyudmil.dakovski@epu.bg

Rumen Trifonov
Department of Computer Systems
Technical University of Sofia
Sofia, Bulgaria
r_trifonov@tu-sofia.bg

***Abstract*—This paper studies the behavior of cooperation between Intelligent Virtual Agents (IVA). The significant role of social emotions and empathy in realizing this complex type of social behavior has been proved. For the purposes of the experiment, a programming system with a scenario has been proposed, according to which intelligent virtual agents with PRE-ThINK architecture sort out the tasks at their workplace. When they mutually show consideration for the plans of the other agent, they manage to realize these plans. Their overall emotional state improves. They are friendly and perfect themselves. The dynamics of making decisions has also been shown: in problematic situations when mixed emotions occur; in justification the choices of tasks they make; in defining priorities and planning actions. Experiments with the agents have been discussed both for the case of collaborative behavior and for lack of communication, comments on the sequences from the chosen types of behavior have been given. The cooperative behavior is achieved by controlling the thoughts of the agents.**

Keywords-Intelligent Virtual Agents (IVA); architectures for IVA; social behavior; mixed emotions; cooperation.

I. INTRODUCTION

One of the biggest challenges in social and biological sciences is to understand the basic and the auxiliary mechanisms, facilitating and favoring the process of collaboration between people and groups of people [4][7]-[10][12][13][15][17]-[19][23][28].

The experiment conducted by Burton-Chellew et al. [23] by means of a public goods game proves that when people unite in groups in order to compete with other similar groups of people for a monetary award, they are more inclined to make greater investments within their own group than when they play individually. The members of the group see themselves as collaborators, not as competitors. The bigger their contribution to the purpose of cooperation, the stronger this image and vice versa. They proved that the strength of the emerging emotions of anger and guilt in an individual, who is a member of a group for cooperation, is a function of both his/her own contribution to the purpose of cooperation, and of the other members' contribution to the purpose of this cooperation.

Cooperation and competition between groups of people can be observed in academic research teams, in the army, in sports teams, etc. [28]. On the one hand, the competition between groups of people is the biggest form of competition, existing in the world. However, on the other hand, the establishment of competing groups favors the arousal of stable and strong cooperation between the group members [17]-[19][23]. This fact has not been studied yet to the depth at which other forms of achieving and maintaining cooperation have already been studied, e.g., by receiving penalties or rewards [6][11][16].

The present paper studies the behavior of cooperation between IVAs with PRE-ThINK architecture. A programming system and a scenario have been proposed, by means of which the complexity of cooperative behavior has been proved. This behavior is regarded here as a social type of behavior, requiring cooperation between the IVAs: sharing; empathy; expressing and understanding social emotions; knowledge of the circumstances, in which a conflict situation could begin; capabilities for recognition of the probable conflict situations; knowledge of the actions for solving or preventing conflict situations. This complex social behavior aims at finding the best solution for all members of a team in every situation. It can prevent from lots of conflict situations; can improve the fulfillment of the tasks; can improve the emotional state of the team members. Therefore, it is expected that the IVAs, which cooperate with the users, will easily gain their trust and will become irreplaceable friends and collaborators in any group united by common interests. For comparison, by means of a separate experiment it has been shown how the lack of cooperation (in particular – lack of communication and empathy) leads to: occurrence of conflict situations and problems; impossibility for the IVAs to achieve their goals and fulfill their plans: worsening the emotional state of the IVAs.

The rest of the paper is structured as it follows: in Section II, the modern tendencies in this field are considered, together with the motivation to model social behavior of collaboration. In Section III, the programming system and the scenario are presented. The concept and nature of the experiments with the system are explained in Section IV. The dynamics of the process of decision making by the IVAs both in case of lack of communication between them and in case of cooperation are considered in Section V. The experimental results are discussed in Section VI. In the Section VII, a number of conclusions are drawn and

 ISBN: 978-1-61208-390-2

directions for further development of the programming system are given.

II. BACKGROUND

Many researchers [3][25]-[27][29] model IVA's behavior in order to establish trust between the user and the IVA. For this purpose IVAs are modeled, having the capability to express so-called moral emotions (regret, joy, compassion, remorse) [5][25]. There has already been much work that promotes cooperation, through trust- and reputation-building models, in multi-agent systems [5].

A great amount of contemporary neurophysiologic research confirms the main role of emotions in rational behavior [1].

Ortony, Clore, and Collins (OCC) model defines a cognitive approach for looking at emotions [30]. This theory is extremely useful for the project of modelling agents which can experience emotions. The cornerstone of their analysis is that emotions are "valence reactions." The authors do not describe events in a way that will cause emotions, but rather, emotions can occur as a result of how people understand events.

The first modification to the OCC model is to allow the definition of different emotions with respect to others, which are known as social emotions. Social emotions can be defined as one's emotions projecting on or affected by others.

According to Lee et al. [24] mixed emotions, especially those in conflict, sway agent decisions and result in dramatic changes in social scenarios. However, the emotion models and architectures for virtual agents are not yet advanced enough to be imbued with coexisting emotions [24]. Modern cognitive architectures, which could be appropriate and would have good results in modeling complex social behaviour are ACT-R [31], Soar [32], CLARION [33], PRE-ThINK [21].

The PRE-ThINK architecture [20][21][22] allows for modelling an IVA, having capabilities to detect and analyze conflicts. Problem situations evoke conflicting thoughts, accompanied by mixed emotions and they are related to a number of different ways of action. The agent considers in advance (Pre-Think) in what way each possible action in a critical situation would reflect over all individuals concerned by it. The originated thoughts are assessed from emotional, rational and needs-related points of view in accordance with the knowledge, priorities and principles of the agent. Agent's behavior motivators are its needs according to Maslow's theory [2].

It is assumed that an IVA, capable of detecting a critical situation, of analyzing it and choosing the best possible option to take care of all individuals concerned, would easily gain trust. Such a behavioral model is presented in this paper with the help of the PRE-ThINK architecture.

III. DESCRIPTION OF THE PROPOSED PROGRAMMING SYSTEM AND SCENARIO

For the purpose of the experiment, a prototype of the programming system and IVAs with PRE-ThINK architecture were modeled. Their structural scheme is given in Fig. 1. The main modules of the programming system are: Module for simulating the passage of time; Module for initialization of the scenario; Module for realization of the scenario; Module for modeling IVA with PRE-ThINK architecture; Module for generation and choice of the thoughts, which will take part in considering the possible

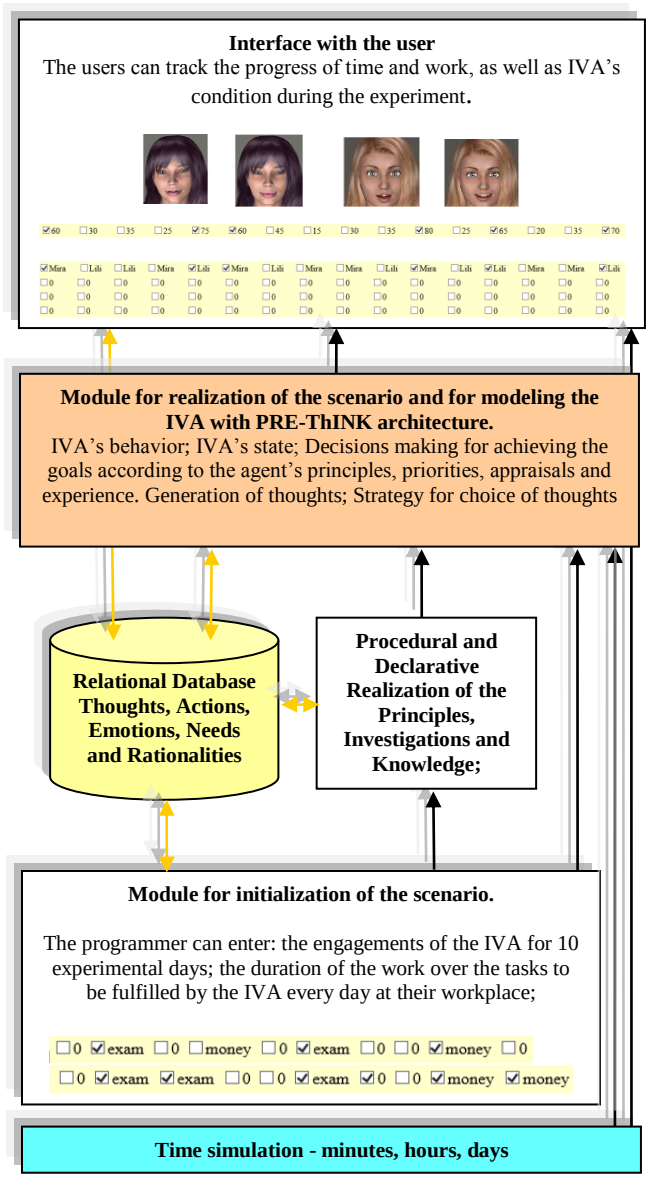


Figure 1. Structural scheme of the program system

actions for solving the conflict situation; Module for decision making; Interface with the user.

The system supports a Relational Database and a Knowledge Base. The components of the PRE-ThINK architecture - Principles, Investigations and Knowledge - are realized in a procedural and declarative way, while the components Thoughts, Actions, Emotions, Needs and Rationalities are presented in relational databases by means of tables and relationships among them. This way of

 ISBN: 978-1-61208-390-2

presenting allows for easy adding or deleting: actions for solving problem situations; thoughts to each possible action; arguments to each thought; emotional, needs-related and rational assessments of each thought-action.

The system allows for applying various strategies for choice of thoughts, which will take part in the process of decision making. For instance: it is possible that all thoughts are taken into consideration in making a decision; it is also possible to define randomly which and how many thoughts will be taken into account; another option is to choose the thoughts depending on the correspondence between their emotional assessment and the emotional state of the IVA or between their needs-related assessment and the needs of the agent at the moment, etc. There are thoughts, deserving attention; there are also thoughts to be suppressed and even overcome (misleading thoughts) in order to come to a good decision. In the experiment, presented here, the thoughts are randomly chosen.

The scenario, proposed and realized especially for the purposes of the present study, includes only participation of IVAs. It is intended that a further development of the programming system and the scenario will also include active participation in the experiment on the side of the users, who will interact with the IVAs and receive advice from them.

According to the proposed scenario for the first part of the study, two IVAs are students and at the same time work for a company. They go to work every day and have a set of tasks to fulfill for the day. The tasks are of various levels of complexity. Those of them, which require more than 45 minutes of work on them, are considered to be complex. Those, which can be fulfilled for 15 to 45 minutes are regarded as light.

It is considered that an IVA has fulfilled his/her daily obligations if he/she has managed to fulfill eight (8) tasks (regardless of their complexity).

An IVA gets a bonus if he/she manages to fulfill four (4) complex tasks within a day.

TABLE I. COMPLEXITY AND DURATION OF THE WORK OVER THE TASKS FOR A DAY AT WORK.

Complexity and Duration	C 60	L 30	L 35	L 25
	C 75	C 60	L 45	L 15
	L 30	L 35	C 80	L 25
	C 65	L 20	L 35	C 70

Every day the IVAs receive exactly sixteen (16) tasks with different time for fulfillment. The number of complex tasks is always six (6). So, the IVAs always have the necessary number of tasks needed for a successful day at work. However, they do not have a sufficient number of complex tasks to be able to get a bonus at the same time. On the other hand, the number of lighter tasks is insufficient for giving both of them the chance to fulfill their obligations by working on them only.

Table I presents an exemplary distribution of the tasks, got by the IVAs at their workplace within a workday. By the symbol C the complex tasks are marked, and the lighter ones are marked by L. The duration of the work over the tasks is measured in minutes.

TABLE II. A) PLANNED ENGAGEMENTS AND INTENTIONS OF THE IVA MIRA FOR THE NEXT 10 DAYS.

Monday	**Tuesday**	**Wednesday**	**Thursday**	**Friday**
	Examination		Party	
Work	Work	Work	Work	Work

Monday	**Tuesday**	**Wednesday**	**Thursday**	**Friday**
Examination			Work for a bonus	Work for a bonus
Work	Work	Work	Work	Work

TABLE II. B) PLANNED ENGAGEMENTS AND INTENTIONS OF THE IVA LILLY FOR THE NEXT 10 DAYS.

Monday	**Tuesday**	**Wednesday**	**Thursday**	**Friday**
	Examination		Work for a bonus	
Work	Work	Work	Work	Work

Monday	**Tuesday**	**Wednesday**	**Thursday**	**Friday**
Party			Work for a bonus	
Work	Work	Work	Work	Work

According to the scenario, the IVAs (Mira and Lilly) sometimes have examinations, sometimes they struggle for a bonus, and sometimes hurry for a party. When they have an exam or hurry for a party, they strive to take 8 light (minor tasks) and leave work as soon as possible. When the IVAs need more money or when they struggle to get a bonus, they strive to fulfill first the required four (4) complex tasks for the day and are ready to stay longer hours at their workplace.

When the IVAs do not have any other engagements for the day except for being at the workplace, and when they do not struggle for a bonus, they fulfill the tasks in their sequence and without any interest to the level of complexity.

TABLE III. PAYOFF MATRIX OF THE PROPOSED SCENARIO

		IVA Lilly	
		Lighter Tasks	*Complex Tasks*
IVA Mira	*Lighter Tasks*	Lack of enough lighter tasks	IVA Lilly - bonus IVA Mira - time
	Complex Tasks	IVA Lilly - time IVA Mira - bonus	Lack of enough complex tasks

For the IVAs, it is important to always fulfill the compulsory eight (8) tasks per day. Therefore, two conflict situations are possible (Table III) – when:

- There is an exam and a party for the IVAs on the same day – (lack of enough lighter tasks);
- Both IVAs have decided to work for a bonus on the same day – (lack of enough complex tasks).

These conflict situations can evoke mixed emotions in the virtual agents. The IVAs have to be able to recognize them and, after consideration, find the best possible solution.

 ISBN: 978-1-61208-390-2

Table II presents all planned engagements of each of the two IVAs (Mira and Lilly) for 10 days.

IV. CONCEPT AND NATURE OF THE EXPERIMENTS

Two experiments with the programming system have been conducted. Each of them tracks the behavior of the virtual agents within ten (10) experimental days.

The following aspects are tracked in each of the two experiments: how many tasks each of the IVAs has fulfilled on each of the considered workdays; what is the level of complexity of the fulfilled tasks; what is the emotional state of each of the virtual agents; what level of importance has been assigned to the different engagements of an agent per day. The aim is to show in which case best results are achieved in terms of work done at achieved personal aims by the IVAs, as well as to show the emotional state of the virtual agents.

In the first experiment, each virtual agent tries to realize his/her aims and engagements for each of the considered workdays not interested in the intentions and aims of the other agent. There is no communication between the two agents. There is no coordinated planning of aims and intentions, correspondingly.

In the second experiment, the virtual agents cooperate between each other. Each of them is interested not only in his/her own engagements and aspirations but also in the aims and intentions of the other agent. At the beginning of the week (on Monday) the agents share and coordinate their engagements for the week (from Tuesday to next Monday). If they think that there may arise a conflict situation on one of the next days, they have two options: 1. To shift the time of their commitments so that the conflict situation is avoided; 2. To choose a day to fulfill in advance some of the complex tasks, planned for days when they are engaged with urgent commitments outside the company.

V. DYNAMICS OF THE PROCESS OF DECISION MAKING

For the purposes of the experiment, IVAs with PRE-ThINK architecture are used [20][21]. The PRE-ThINK architecture consists of the following components: Principles, Rationalities, (+/-) Emotions, Thoughts, Investigations, Needs and Knowledge.

The IVA makes his/her decisions based on his/her principles. The following IVA principles have been modelled: “Choose the better possible action"; "Neglect the basic needs until reaching a definite threshold of dissatisfaction, giving priority to the highest-order needs”; “Evaluate the desires and commitments of the other IVAs as your own”; “Your personal commitments are as important as those, of the others”.

The agent has a set of thoughts related to the fulfillment of the tasks at the workplace and outside the company – thoughts, related to: self-observation; the other IVA’s state; the way in which the agent could help to solve the problematic situations.

Each thought is related both to an emotion, and to a need, and also has its rational component – importance – with a value from 1-3. Each thought is also related to an action.

As it concerns the emotional component of the thoughts – only the following emotions are taken into account: anxiety – a negative emotion with a value of (-1) and gladness – a positive emotion with a value of (+1).

The hierarchy of needs according to Maslow’s theory [2] is used – physiological (ph), safety (s), love and belonging (lb), esteem and self-assessment (es), self-actualization (sa), aesthetics (a). Weights of the needs are introduced – Wneed, corresponding to their priority: Wph,=10; Ws=20; Wlb=30; Wes=40; Wsa=50; Wa=60; When, because of the occurrence of an event, one or more needs prove to be unfulfilled, i.e., there is a crisis situation, then the needs rearrange so that the unfulfilled ones receive first priority. The unfulfilled needs are arranged in an order, opposite to the order of needs weights in a normal state of the agent.

Each action of the IVA is related to a need: the performance of the duties – to the need of safety; the meetings with friends – to the need of love and belonging; the struggling for a bonus – to the need of esteem and self-esteem; the examinations – to the need of self-actualization; the possession of more funds – to the aesthetics needs.

Let a thought addressed to the situation s be denoted by Th_s. If the importance of the thought Th_s is denoted by $I_{imp.Th\_s}$, the weight of the need, related to this thought is expressed by W_{needTh_s}, the emotion implied by this thought is marked by $E_{emot.Th\_s}$, then, following the formulae for calculating the assessment value of the thought A_{Th_s}, corresponding to the situation s, will be [21]:

$$A_{Th\_s} = E_{emot.Th\_s} * W_{needTh\_s} * I_{imp.Th\_s} \quad (1)$$

If a thought is partially related to more than one need, then the sum of the weight percentages of the needs to which it is related is taken into account in the formulae.

Each thought is related to an action. The assessment values of the thoughts related to one and the same action in one and the same situation are put on the one basin of the “thoughts balance”. The assessment values of the thoughts for the same situation, but related to another action, are put on another basin etc. Our “thoughts balance” will have as many basins as the alternative actions considered by the agent in the particular situation are. The module of the assessment values is summed and the action from the basin having the highest assessment value is undertaken [21].

A. The First Experiment

Here is an example of the thoughts, generated during the first experiment, when both IVAs only think about their own planned actions and are not interested in the plan of the other IVA. The second day of the experiment, when the IVAs Mira and Lilly have an examination is considered.

Mira’s first thought:

I have an examination today so I will fulfill 8 easy tasks in order to leave work at the earliest and go to the exam.

A thought, focused on the fulfillment of the tasks at the workplace and on the examination. Positive emotion – gladness $E_{emot.Th1\_1} = 1$, that she will manage both to pass the exam and to fulfill the tasks; rational component – importance with value $I_{imp.Th1\_1}=3$; motivator – the need of

self-actualization Wsa=50 and safety Ws=20, $W_{needTh1\_1}$=70, *action – going to the examination after fulfilling the compulsory 8 (easy) tasks for the day.*

$$A_{Th1\_1(work\_exam)} = 1*70*3 = 210 \quad (2)$$

Mira's second thought:

I have an examination today. However, I have to be at work and therefore I will not go to the exam.

A thought, focused on the fulfillment of the tasks at the workplace. Negative emotion – anxiety about the safety at the workplace $E_{emot.Th2\_1}$ *= -1; rational component* $I_{imp.Th2\_1}$ *= 1; motivator – the need of safety with weight Ws=20;* W_{needTh_s} *= 20; action – postpones the examination.*

$$A_{Th2\_1(work)} = -1*20*1 = -20 \quad (3)$$

Mira's third thought:

I have an examination today and therefore I will not go to work.

A thought, focused on the examination. Negative emotion – anxiety about the examination - $E_{emot.Th3\_1}$ *= -1; rational component –* $I_{imp.Th3\_1}$ *=1; motivator – the need of self-actualization Wsa=50;* $W_{needTh3\_1}$ = 50*; action – goes to the examination and does not go to work.*

$$A_{Th3\_1(exam)} = -1*50*1 = -50 \quad (4)$$

The thoughts about the two alternative actions are weighed as if on a balance and the IVA takes the decision for action.

Thoughts of going to the exam:

$A_{Th1\_1\ (work\ and\ exam)}$ = 210

$A_{Th3\_1\ (exam)}$ = -50

Thoughts of postponing the exam:

$A_{Th2\_1\ (exam)}$ = -20

It is obvious that here the thoughts of going to the exam outweigh. The most important thought which will be realized is the first thought. The IVA Mira will go to work in order to fulfill the norm of 8 (though easy) tasks and then will go to the exam in time.

B. The Second Experiment

In the second experiment, when the IVAs consider the first conflict situation, the following thoughts are generated:

Mira's first thought:

Lilly and I have an examination on the same day. i.e., we have the same commitment with the same priority. Consequently, it will not be fair if any of us gives up her commitment.

A thought, focused on the relationship between the agents; negative emotion – anxiety about the exam $E_{emot.Th1\_2}$ *=-1; rational component – importance with a value of* $I_{imp.Th1\_2}$ *=3; motivator – need of love and belonging Wlb=30;* $W_{needTh1\_2}$=30; *action – both of them go to the exam without postponing it.*

$$A_{Th1\_2(agent-exam)} = -1*30*3 = -90 \quad (5)$$

Mira's second thought:

Lilly and I have exams on the same day. We could postpone our exams in order to fulfill our task at the workplace by a schedule. There will be next dates for these exams.

A thought, focused on the relationships between the agents and the priorities at the workplace; negative emotion – anxiety about the exam $E_{emot.Th2\_2}$ *=-1; rational component – importance with value* I_{impTh2_2} *= 2; motivator – the need of safety at the workplace Ws=20;* $W_{needTh2\_2}$=20; *action – fulfilling the obligations at the workplace by a schedule and postponing the exams.*

$$A_{Th2\_2(agent-work)} = -1*20*2 = -40 \quad (6)$$

Mira's third thought:

If we go to our exams tomorrow, we could take and fulfill today two of the complex tasks, envisaged for tomorrow. Thus we will be able to follow the work schedule on the one hand, and go to the exam, on the other hand.

A thought, focused on the relationships between the agents at the workplace, on the work and on the exam; positive emotion – safety and gladness - $E_{emot.Th3\_2}$ *=1; rational component – importance with value* $I_{imp.Th3\_2}$ *= 3; motivators – the need of safety with weight Ws=20 and the need of self-actualization with weight Wsa=50;* $W_{needTh3\_2}$ = 70; *action – fulfilling the complex work tasks in advance (since there will not be any time to fulfill them on the day of the exam), going to the exam without postponing it.*

$$A_{Th3\_2(agent-exam-work)} = 1*70*3 = 210 \quad (7)$$

Thoughts in support of the idea of both IVAs going to the exams:

$A_{Th1\_2\ (agent\_exam)}$ = -90

$A_{Th3\_2\ (agent\_exam\_work)}$ = 210

Thoughts in support of the idea of both IVAs postponing their exams:

$A_{Th2\_2\ (agents\_work)}$ = - 40

The thoughts about the two alternative actions are weighed as if on a balance and the IVA takes the decision for action.

It is obvious that here overweigh the thoughts of Mira and Lilly of going to their exams after fulfilling in advance the complex tasks, for which there will be no time on the day of the exams.

One more example:

In the second experiment, when the IVA Lilly has planned a party and Mira has an examination, Lilly's thoughts are the following:

Thought 1:

I have planned a party for Monday, but Mira has an examination. She would probably also like to leave work earlier. The examination is more important than a party and I could help her by organizing the party on the next day. We will all be glad and have a good time together. We will also fulfill our tasks at the workplace.

A thought focused mainly on the relationship between the agents; positive emotion – gladness from the opportunity to

help in solving the complex situation $E_{emot.Th1\_3}$ *=1; rational component – importance with value* $I_{imp.Th1\_3}$ *= 3; motivator – the need of love and belonging with weigh Wlb=30;* W_{needTh_s} = 30; *action – having a party on the next day, which is free of commitments.*

$$A_{Th1\_3(agent-work)} = 1*30*3 = 90 \quad (8)$$

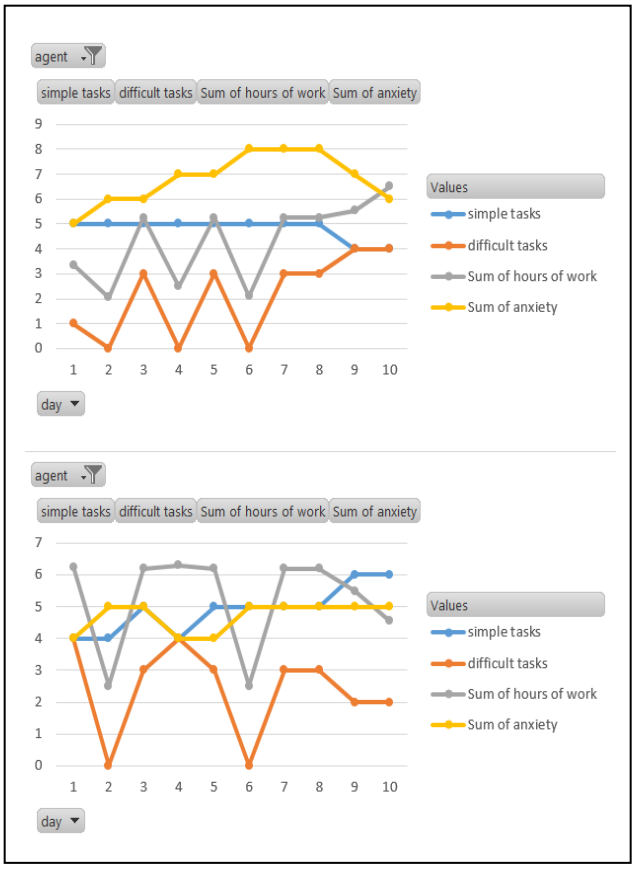


Figure 2a. Experimental results for the case of no cooperation between the IVAs (Mira and Lilly). It can be seen that there are no complex tasks fulfilled on the second and on the sixth day. On the nineth and tenth day only IVA Mira fulfills the needed 4 complex tasks in order to receive a bonus. Anxiety in the IVAs stays high.

Thought 2:

If I do not postpone the party and Mira does not postpone the exam, we will not be able to fulfill our tasks at the workplace. I could try to work overtime, but I will be very tired.

A thought, focused mainly on the fulfillment of the obligations at the workplace; negative emotion – regret $E_{emot.Th2\_3}$ *=-1; rational component – importance* $I_{imp.Th2\_3}$ *= 1; motivator – need of safety with weight Ws=20;* $W_{needTh2\_s3}$ =20; *action – work over the complex tasks in advance.*

$$A_{Th2\_3(work)} = -1*20*1 = -20 \quad (9)$$

The thought to postpone the party for the next free of commitments day outweighs here obviously and this is the action, which is taken up.

VI. Discussion of Experimental Results

The results from the described above experiments with the IVAs Lilly and Mira are given in Fig. 2a and Fig. 2b. For each of the ten (10) observed days the following data are shown: the number and the level of complexity of the fulfilled tasks; the duration of work of the agents, measured in hours; the summarized emotional state of the IVA depending on whether they have managed to realize all their stated commitments.

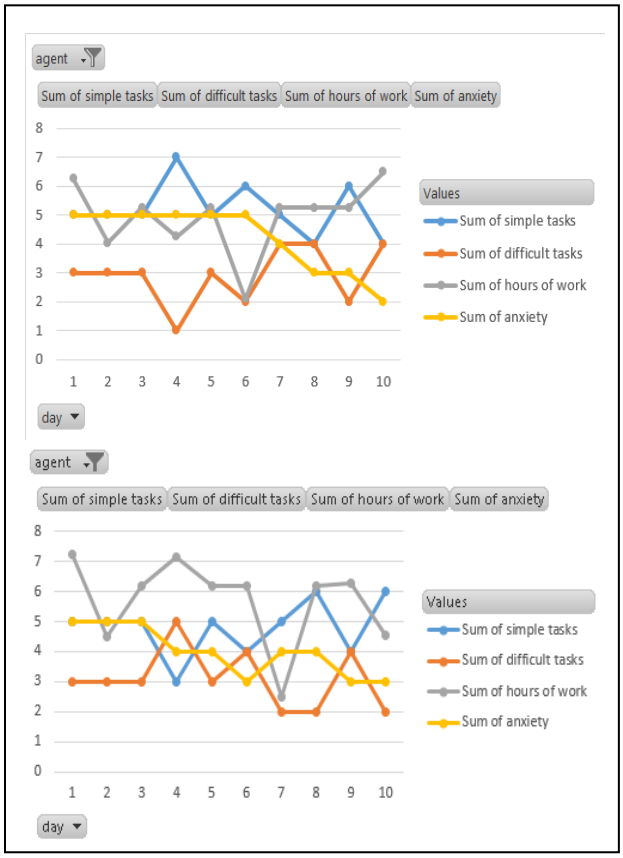


Figure 2b. Experimental results for the case when the IVAs distribute their commitments in advance. (Mira and Lilly). It can be seen that on the second and on the sixth day all complex tasks are fulfilled. Mira manages to fulfill the needed 4 complex tasks and receives a bonus on the nineth day of the expeiment, while Lilly receives her bonus on the tenth day. The tendency is that anxiety reduces with time due to averting the conflict situations.

From Fig. 2a, it can be seen that, in case of no communication, the IVAs often do not manage to fulfill their work obligations because they have other tasks with greater priority. For instance, during the first week on Tuesday and the second week on Monday, they leave work without fulfilling their norm because they have examinations and meetings with friends. On the other hand, on Thursday the second week they try to work more and earn a bonus but fail again. It turns out that because of the lack of communication and coordination between them, they compete for the complex tasks for the day.

Thus, day by day, their anxiety and dissatisfaction with respect to the workplace grow. If this tendency continues a

longer time the threshold of dissatisfaction for the need of safety will be overcome.

When this happens, the fulfillment of the obligations at the workplace will become first priority for the IVAs. All other commitments will have lower priorities. In further conflict situations they will have to miss examinations or meetings with friends. Their anxiety and dissatisfaction will grow further. Then, the IVAs may face extremely complex problematic situations and they will have to choose what to give up – work, university or friends. They may have to look for a new job or a new subject to study at the University. The occurrence of these extremely complex situations can be prevented by improving the communication and by cooperation between the agents.

From Fig. 2b it can be seen that the IVAs have managed to redistribute their plans and commitments in a way that has allowed them to prevent all conflict situations.

Firstly: Lilly has decided to meet her friends on Tuesday – the 7th day of the experiment – instead on Monday – the 6th day of the experiment. Thus on Monday Mira will be able to fulfill most of the easier tasks and go to her exam in time. Secondly: Mira has decided on Wednesday (the 8th day of the experiment) to work for a bonus, instead on Thursday (9th day of the experiment). Thus she gets a bonus, and Lilly gets her bonus on the next day (9th) of the experiment.

The most complex conflict occurs on the second day (Tuesday, first week) of the experiment, when both IVAs have examinations. This is because the assessment for importance of the commitment is one and the same. It is impossible to make a decision whose exam to be postponed. Therefore, they decide that each of them will fulfill two of the complex tasks on the day before the exams. It means that on the day of the exams they will have to fulfill five (5) easier tasks and one (1) complex task each. Thus only by coordinating their work, the IVAs manage to achieve all their goals and commitments and there is no stress at their workplace. On the contrary, they are glad because they get bonuses whenever they want, pass their exams and meet their friends with no problems.

Of course if the needs of money grow more or if the commitments outside the company grow more, this solution will not be good enough. But in the cases when the possibilities of coordination and cooperation between the IVAs are exhausted, the solution can be found:

1) By employing more IVAs (for the cases when Mira and Lilly will have to leave work earlier), and

2) By increasing the number of complex tasks, available for the two IVAs (for the cases when they more often want to work for bonuses).

VII. CONCLUSION

Cooperation is a complex social behavior, requiring empathy and preliminary consideration. It is related to planning and making decisions in complex situations, when mixed emotions arise. The realization of cooperation requires understanding of behavior, emotions, reasons and interests both of us and of the other participants in a considered scenario. This paper studies the cooperative behavior between IVAs. The significant role of social emotions and empathy in realizing this complex type of social behavior has been proved. For the purposes of the experiment, a programming system with a scenario have been proposed, according to which intelligent virtual agents with PRE-ThINK architecture distribute between the two of them the tasks at their workplace. When they mutually show consideration for the plans of the other agent, they manage to realize these plans. Their overall emotional state improves. They are friendly and self-actualize themselves. The dynamics of making decisions has also been shown as it follows: in problematic situations when mixed emotions occur; in justification the choice of tasks to fulfill; in defining priorities and planning actions. Experiments with the agents have been discussed both for the case of cooperative behavior and for lack of communication, comments on the sequences from the chosen types of behavior have been given.

Based on his/her own principles, knowledge and priorities in a critical situation, the agent evaluates the possibilities for action from emotional, rational and needs-related point of view and chooses the best possible action. The purpose of the software agent is to possibly take the best care of all his collaborators. It is assumed that such behavior would facilitate the establishment of trust between the IVAs and the users on the one hand; on the other hand, it could avert a great part of the conflict situations, which occur every day, including at the workplace (mainly due to the lack of cooperation and empathy).

It is envisaged to develop this prototype of a programming system by a learning module. It will allow the IVAs to learn by assessing not only their thoughts but also the thoughts of the other IVAs, with which they cooperate. When an IVA decides that another IVA's thought is a sufficiently strong argument in favor of a given action, then he/she will be able to save and use it in the future.

Thus, through communication, the IVAs will enrich with new and stronger arguments in support of a given action; they will gain more and more trust; they will be more and more useful for the users and will become irreplaceable, precious friends and members of every team in the world and in each sphere of life.

Experiments, in which the different IVAs will use different strategies for choosing the thoughts to take part in considering the possible actions for solving a given problematic situation, will be of interest.

The experiments will be extended with the aim to cover situations, in which rearrangement of the IVA's priorities will occur. An interesting question will arise in relation to the way in which the IVAs express empathy when their priorities are in a different order. The assessments of the desires and aims of the others will be different then. Certain arguments will look value for some IVAs, other arguments will be important to other IVAs. We believe that the results from such experiments will be of great interest to the scientific world.

It is intended that the programming system is extended by allowing the users to participate in the scenario, to communicate with the IVAs and receive advice from them.

 ISBN: 978-1-61208-390-2

The programming system can be useful as a Socially Assistive Application (SAA). In general, SAA are intended to motivate the users and make them change their social behavior. These applications are useful not only for people with social deficits, since all people sometimes need to share a problem or an experience and receive advice from a friend or a specialist. On the other hand, Cloud computing technologies give the chance to realize in the cloud data centers, knowledge base, task planners, deep learning, information processing. This will allow the users, IVAs and robots to share knowledge about and solutions for problematic situations and apply them when necessary.

And last but not least, modeling cognitive processes will help for their better understanding and management. This, in its turn, will lead to a better quality of life.

ACKNOWLEDGMENT

This research is supported by: the Technical University of Sofia and its Branch in Plovdiv; the Centre for education and innovation at TU-Sofia; and the Scientific and research sector at TU-Sofia.

REFERENCES

[1] A. R. Damasio, "Descartes' Error. Emotion, Reason and the Human Brain," Avon Books, 1994.

[2] A. H. Maslow, "Motivation and personality," Addison Wesley Longman, Inc. USA, 1970.

[3] R. S. Picard, "Affective computing," The Mit Press, Cambridge, Massachusetts, London. 1998.

Article in a journal:

[4] S. Bowles and H. Gintis, "The evolution of strong reciprocity: cooperation in heterogeneous populations. Theoretical Population," Biology, 65(1), 2004, pp. 17−28.

[5] C. M. deMelo, P. Carnevale, and J. Gratch, "The Impact of Emotion Displays in Embodied Agents on Emergence of Cooperation with People," Teleoperators and Virtual Environments, vol. 20, no. 5, 2012, pp. 449-465.

[6] E. Fehr and S. Gachter, "Altruistic punishment in humans," Nature, 415(6868), 2002, pp. 137−140.

[7] H. Gintis, "Strong reciprocity and human sociality," Theoretical Biology, 206(2), 2002, pp. 169−179.

[8] H. Gintis, "The hitchhiker's guide to altruism: Gene-culture coevolution, and the internalization of norms," Theoretical Biology, 220(4), 2003, pp. 407−418.

[9] H. Gintis, S. Bowles, R. Boyd and E. Fehr, "Moral sentiments and material interests," The foundations of cooperation in economic life., Cambridge, MA: MIT Press, 2005.

[10] A. Gunnthorsdottir and A. Rapoport, "Embedding social dilemmas in intergroup competition reduces free-riding, Organizational Behavior and Human Decision Processes," 101(2), 2006, pp. 184−199.

[11] O. Gurerk, B. Irlenbusch, and B. Rockenbach, "The competitive advantage of sanctioning institutions," Science, 312, doi:10.1126/science.1123633, 2006, pp. 108–111.

[12] J. Henrich, R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis, R. McElreath, M. Alvard, A. Barr, J. Ensminger, M. S. Henrich, K. Hill, F. Gil-White, M. Gurven, F.W. Marlowe, J. Q. Patton, and D. Tracer, "Economic man in cross-cultural perspective: behavioral experiments in 15 small-scale societies," The Behavioral and Brain Sciences, 28 (6), [discussion 815–55], 2005, pp. 795−815.

[13] L. Lehmann, F. Rousset, D. Roze and L. Keller, "Strong reciprocity or strong ferocity a population genetic view of the evolution of altruistic punishment," Vol 170, pg 21, American Naturalist, 2007, 170(4):661.

[14] J. Sabater, C. Sierra, "Review on computational trust and reputation models," Artificial Intelligence Review, 2005, pp. 33—60.

[15] J. H. W. Tan and F. Bolle, "Team competition and the public goods game," Economics Letters, 96(1), 2007, pp. 133−139.

[16] C. Wedekind and M. Milinski, "Cooperation through image scoring in humans," Science, 288(5467), 2000, pp. 850−852.

[17] S. A. West, A. S. Griffin, and A. Gardner, "Evolutionary explanations for cooperation," Current Biology, 2007a, 17(16), R661−R672.

[18] S. A. West, A. S Griffin and A. Gardner, "Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection," Evolutionary Biology, 2007b, 20(2), pp. 415−432.

Article in a conference proceedings:

[19] R. Boyd, H. Gintis, S. Bowles and P. J. Richerson, "The evolution of altruistic punishment," The National Academy of Sciences of the United States of America, 100(6), 2003, pp. 3531−3535.

[20] D. Budakova, L. Dakovski, "Computer Model of Emotional Agents," IVA'06, LNAI 4133, poster, 2006, pp. 450.

[21] D. Budakova, "Behavior of Home Care Intelligent Virtual Agent with PREThINK Architecture," ICAART'2011, Rome, Italy, 2011, pp. 157-167.

[22] D. Budakova, L. Dakovski, "Social Behavior investigation of An Intelligent Virtual Agent with the help of typical working student's life scenario modeling," ICAART 2012, Vilamoura, Algarpe, Portugal, 6-8 February, ISBN: 978-989-8425-96-6, 2012, pp. 317-324.

[23] M. N. Burton-Chellew, A. Ross-Gillespieb, S. A. West, "Cooperation in humans: competition between groups and proximate emotions, Evolution and Human Behavior", Elsevier, 31, 2010, pp. 104–108.

[24] B. P. H. Lee, E. Ch. Ch. Kao, V. W. Soo, "Feeling Ambivalent: A Model of Mixed Emotions for Virtual Agents," IVA 2006, LNAI 4133, Springer, 2006, pp. 329-342.

[25] C. M. de Melo, L. Zheng and J. Gratch, "Expression of Moral Emotions in Cooperating Agents, Intelligent Virtual Agents," Amsterdam, Sep. 2009, pp. 14-16.

[26] J. Gratch, N. Wang, J. Gerten, E. Fast and R. Duffy, "Creating Rapport with Virtual Agents," International Conference on Intelligent Virtual Agents, Paris, France, 2007.

[27] R. Niewiadomski, M. Ochs and C. Pelachaud, "Expressions of empathy in ECAs," IVA'2008, Tokyo, Japan, LNAI, Vol. 5208, 2008, pp. 37-44.

[28] M. Puurtinen and T. Mappes, "Between-group competition and human cooperation," Royal Society B-Biological Sciences, 276(1655), 2009, pp.355−360.

[29] T. W. Bickmore, L. M. Pfeifer and M. K. Paasche-Orlow, "Health Document Explanation by Virtual Agents," IVA'07, Paris, France, LNAI, 2007, pp. 183-196.

[30] A. Ortonty, G. L. Clore, A. Collins, "The Cognitive Structure of Emotions," Cambridge University Press, 1988.

[31] C. Dimov, J. N. Marewski and L. J. Schooler, "Constraining ACT-R models of decision strategies: An experimental paradigm," Cognitive Science Society, Austin, 2013, pp. 2201-2206.

[32] S. Mohan, A. H. Mininger, J. E. Laird, "Towards an Indexical Model of Situated Language Comprehension for Real-World Cognitive Agents," Conference on Advances in Cognitive Systems, 2013, pp. 153-170.

[33] R. Sun, "Motivational representations within a computational cognitive architecture," Cognitive Computation, Vol.1, No.1, 2009, pp. 91-103

 ISBN: 978-1-61208-390-2

Directional-Change Event Trading Strategy: Profit-Maximizing Learning Strategy

Monira Essa Aloud
Department of Management Information Systems, College of Business Administration
King Saud University, KSA
email: mealoud@ksu.edu.sa

***Abstract*—Many investors seek a trading strategy in order to maximize their profit. In the light of this, this paper derived a new trading strategy (DCT1) based on the Zero-Intelligence Directional Change Trading Strategy ZI-DCT0, and found that the resulting strategy outperforms the original one. We enhanced the conventional ZI-DCT0 by learning the size and direction of periodic fixed patterns from the price history for EUR/USD currency pairs. To evaluate DCT1, experiments were carried out using the bid and ask prices for EUR/USD currency pairs from the OANDA trading platform over the year 2008. We compared the resulting profits from ZI-DCT0 and DCT1. The analysis revealed interesting results and evidence that the proposed DCT1 investment strategy can indeed generate effective electronic trading investment returns for investors with a high rate of return. The results of this study can be used further to develop decision support systems and autonomous trading agent strategies for the FX market.**

Keywords-Trading strategies; Autonomous trading agent strategies; Pattern recognition; FX Market.

I. INTRODUCTION

Electronic trading strategies have become a hot topic in the field of financial markets, and numerous strategies have been developed. Investors are always looking for a trading strategy that maximizes their profits. The financial literature has featured a long debate on the effectiveness of the technical analysis of financial market time series [1–8]. Some argue that prices are not predictable based on historical information, since all the relevant public information is mirrored in the prices. In contrast, recent studies [8,9] have uncovered empirical evidence of various price anomalies, and therefore have confirmed positive observed evidence on the effectiveness of technical analysis for analyzing financial price time series.

Trend Following (TF) trading strategy is a widely used investment strategy due to the simplicity of the principle on which it is based and its effectiveness [10–13]. TF adopts a rule-based investment strategy based on the directions of market price trends, where a trader takes advantages of the price trend on the assumption that the current price trend will continue in the same direction. Furthermore, the underlying assumption of TF is that a trader will follow the price trend with the assumption that some traders have market information prior to the general public which is reflected in the direction of the price trend [10]. A TF trader places a buy order when the price is rising, while a sell order is placed when the price is falling. The financial literature reveals successful investments based on a TF trading strategy in stock markets [11], currency markets [14] and commodity futures' markets [13]. Similar to the TF investment strategy is the Contrary Trading (CT) strategy with regard to the direction of the market price trend. A CT trading rule places a buy order in anticipation that the price will move in the opposite direction. For example, a trading rule may indicate a buy order opportunity when the price falls by 0.03% and afterwards places a sell order if the price rises by 0.06%.

Despite the effectiveness of TF and CT investment strategies, comparatively few works have explored the application of learning in order to enhance TF and CF investment strategies. Aloud et al. [15] have constructed a trading strategy called ZI-DCT0 based on pooling two approaches: (i) the DC event approach [16] and (ii) TF and CT investment approaches. Trading in the financial markets is highly active at some times, but calm down at others which makes the flow of physical time discontinuous. For that reason using fixed time scales for studying the price changes in the market runs the risk of missing important price activities. The DC event approach captures periodic activities in the price time series to detect major periodic patterns based on the trader's expectations of the market. Given a fixed threshold size, the DC approach characterizes periodic price trend movements in the price time series, where any occurrence of a DC event represents a new intrinsic time unit, independent of the notion of physical time change. A comparable trading strategy to the ZI-DCT0 is introduced by Alfi et al. [17] in which a trader places an order if the price fluctuations exceed a defined threshold. The threshold is determined by the trader, and remains constant during the traders' trading period in the market. The main difference between ZI-DCT0 and the one introduced in [17], is that ZI-DCT0 considers the direction and the overshoot of price movements in the traders' trading activities.

In the light of this, the work reported in this paper introduces a new trading strategy called Directional-Change Trading (DCT1) derived from ZI-DCT0, where ZI-DCT0 has been enhanced by the incorporation of a learning model fed by historical dataset, with the aim of determining the size and direction of periodic patterns in the price time series. As such, the DCT1 is able to recognize periodic patterns in a price time series such as DC events. This may be the key to providing effective decision support for traders in the financial markets. DCT1 applies a simple learning mechanism which avoids the complexity of artificial intelligent trading strategies and also the vagueness of zero-intelligence and Buy-and-Hold trading strategies in which traders trade randomly, subject to budget constraints.

The rest of the paper is organized as follows. A brief literature review of related works is presented in Section II. The ZI-DCT0 trading strategy is described in Section III. The new trading strategy is depicted in Section IV together with a description of the core mechanism of the DCT1 trading strategy. The experimental design and results are presented

 ISBN: 978-1-61208-390-2

in Section V. A summary and conclusions are provided in Section VI.

II. Related Work

A considerable amount of scientific research has explored artificial intelligence and cognition techniques in terms of financial data processing and filtering, knowledge discovery and building potential adaptive trading strategies for investment in the financial markets. Agent-based adaptive systems have been applied effectively to study and understand complex financial market phenomena as a means of studying the behaviour of individual agents within a financial market setting, the trading interaction effects, emergent macro properties, knowledge discovery with regard to trading behaviour, designing adaptive investment trading strategies, and the impact of market rules or policies, among others.

Our paper relates to a large body of literature on investment trading strategies that is too vast to survey here. Therefore, we limit our literature review to the most influential works with regard to designing investment trading strategies. In the literature, the design of the trading strategy ranges from simple budget constrained Zero-Intelligence (ZI) strategy as in [18–20], to complicated intelligent strategy, such as in [21,22].

ZI strategy was introduced by Gode and Sunder [20] to examine the continuous double-auction (CDA) mechanism where the strategy implies random trading, subject to budget constraints. Thus, ZI strategy does not carry out observation nor learning the price trend movements. Gode and Sunder's experimental results show that markets operating with human traders and ZI constrained traders converged to the equilibrium price whereas the market operating with ZI unconstrained traders did not. In following work, Gode and Sunder [23] examined the lower-bounds of the level of learning required for a trading strategy to achieve sufficient outcomes. The results show that a simple budget constrained ZI strategy is capable to accomplish satisfactory outcomes. Thus, they conclude that learning is not required. Their results show that the theoretical equilibrium price in financial markets is determined more by market structure than by the level of intelligence of the traders in that market.

Alfi et al. in [17,24,25] show that representation of trading strategy uses to great extend a simple learning mechanism which depends on observing the price's movements. In particular, each agent commits to a fixed threshold, as a result the agent places an order when the price fluctuations are above this threshold. The strategy does not consider the direction and the overshoot of the price movement where the overshoot represents the size of the price movement beyond the defined threshold by the agent.

The availability of historical financial data provides extensive rich resources for predicting future price changes in financial time series. The possible reward of a useful tool for forecasting changes in the price of financial assets is without doubt a major motivator. Financial forecasting has attracted the attention of researchers from a variety of computer science areas. Techniques from Artificial Intelligence, and in particular Evolutionary Computation, have been used extensively in the design of financial forecasting techniques for predicting future price changes in financial time series. The most commonly used techniques are Artificial Neural Networks (ANNs) [26], Genetic Algorithms (GAs) [27], Genetic Programming (GP) [28] and Learning Classifier Systems (LCS) [29,30]. Austin et al. [31] provide an overview of the research conducted by the Centre for Financial Research at Cambridge University's Judge Institute of Management, which has been researching trading techniques in FX markets for forecasting intraday or daily exchange rates.

In this section, our intention is to provide a brief illustrative description of the artificial intelligence techniques used in financial forecasting and provide a brief description of some of the most relevant works in the field.

In financial forecasting, ANNs are probably the most heavily exploited artificial intelligence technique. There are many studies in the literature on the subject of ANNs. Amongst them are [32–35]. The ANNs technique has been applied in diverse areas of finance. Wong and Selvi in [26] provide a good survey of the literature on ANNs between 1990 and 1996. An additional recent survey can be found in [36]. Yao and Tan in [37] provide empirical evidence of the appropriateness of ANNs with regard to the prediction of foreign exchange rates.

In financial forecasting, the input data is a fundamental issue in terms of the success or failure of ANNs, as is the case with other forecasting techniques. Nevertheless, this issue is particularly important in the case of ANNs owing to the lack of flexibility of ANNs techniques. A number of works in which ANNs are combined with genetic algorithm techniques is demonstrated by [38]. Further relevant examples are the studies done by [33–35,39–43].

GAs were invented by John H. Holland in 1975 [27]. GAs belong to the evolutionary algorithms field which offers very popular techniques in optimization and machine learning problems. GAs use a generation of individuals where each individual is a candidate for a possible solution to the problem. A new population is produced by means of selecting the fittest individuals from the current population through the application of genetic operators such as crossover and mutation.

In the finance area, GAs are not limited to forecasting. GAs are not just used for forecasting but they are also important in modelling learning in an ABM. Examples of relevant works which have involved financial forecasting using the GAs technique are [44–48]. There is a number of limitations of GAs, such as the fixed size structure of the individuals and their representation. Nevertheless, GAs can be used as a meta-heuristic or in combination with other forecasting techniques, to advance the predictions' performance as can be seen in the works done in [38,49].

LCS is a machine learning mechanism wherein a population of rules is evolved and modified by genetic algorithms. Since GAs are used to select and modify the population of rules, this means that the representation of the rules has to be done with binary strings. Holland and Miller [50] proposed the use of an LCS to model economic agents. The SF ASM used an LCS to forecast price changes in time series [51]. In addition, an LCS was used to perform financial forecasting in the work

 ISBN: 978-1-61208-390-2

done in [29,30].

LCS has the same limitations as GAs in terms of the representation and the fixed size structure of the individuals of the population. An alternative technique to the GAs and LCS is the genetic programming technique.

The reported studies above adopted different investment strategy methodologies to perform pattern deduction in price financial market time series. However, these methodologies require demanding financial interpretation ability which means that they cannot be used by regular investors in order to interpret actual investment trading behavior.

III. ZI-DCT0

ZI-DCT0 [15] is a trading strategy based on the DC event approach [16]. DC event approach is an approach for studying the financial time series based on intrinsic time rather than physical time. Physical time adopts a point-based system while intrinsic time adopts event-based system. Physical time is homogenous in which time scales equally spaced based on the chosen time unit (e.g., seconds). In contrast, intrinsic time is irregularly-spaced in time given that time triggers at periodic events of price revolution. The basic unit of intrinsic time is an event where event is the total price change exceeding a given fixed threshold defined by the observer.

Given a fixed threshold of size ($\triangle x_{DC}$), the absolute price change between two local minimum and maximum prices is decomposed into directional-change (DC) event of size $\triangle x_{DC}$ and its associated overshoot (OS) event. The OS event is the absolute price change beyond the $\triangle x_{DC}$ threshold. A DC event can be either an upturn event or a downturn event. An upward run is a period between an upturn DC event and the next downturn event. In contrast, a downward run is a period between a downturn DC event and the next upturn DC event. A downturn (upturn) DC event terminates an upward (downward) run, and starts an upward (downward) run.

Prior to studying and analyzing an asset's price time series, two variables are defined: the last high and low prices where they are assign as initial value the asset's price at the start of the price time series sequence. During an upward run, the last high price is continuously adjusted to the maximum of the current price p_t at time t and the last high price. During the period of a downward run, the last low price is continuously adjusted to the minimum of the current price p_t at time t and the last low price. An upturn DC event occurs once the absolute price change between the current price and the last low price is higher than the defined threshold of size $\triangle x_{DC}$. In contrast, a downward DC event occurs once the absolute price change between the current price and the last high price is lower than the defined threshold of size $\triangle x_{DC}$.

A ZI-DCT0 commits himself to a fixed threshold and a method for trading where the method can be one of two forms: CT or TF trading. Algorithm 1 demonstrates the trading mechanism for ZI-DCT0. ZI-DCT0 provide evidence in [52] to generate good quality trading strategy in terms of the trader's return of investments, analyzing price movements and reproducing the statistical properties of the FX market trading behavior when used in an agent-based market. The restrictions of ZI-DCT0 is the randomness and consistent in choosing the threshold and the type of trading.

Algorithm 1 The core trading mechanism for the ZI-DCT0.

Require: initialise variables (event is upturn event, x =p_t, $\triangle x_{DC}$ (Fixed) ≥ 0)
if event is upturn event **then**
  if $p_t \leq x \times (1 - \triangle x_{DC})$
    event $\leftarrow$ downturn event
    x$\leftarrow$ p_t
    Sell$\rightarrow$ZI-DCT0 CT, Buy$\rightarrow$ZI-DCT0 TF
  else
    x$\leftarrow$ max (x, p_t)
  end if
else //Event is downturn event
  if $p_t \geq x \times (1 + \triangle x_{DC})$ **then**
    event $\leftarrow$ Upturn event
    x$\leftarrow$ p_t
    Buy$\rightarrow$ZI-DCT0 CT, Sell$\rightarrow$ZI-DCT0 TF
  else
    x$\leftarrow$ min (x, p_t)
  end if
endif

IV. DCT1

DCT1 is an intelligence trading strategy driven from the ZI-DCT0 in which the strategy involves learning toward identifying the estimated size and the direction of periodic patterns from an asset's price time series. DCT1 aims to overcome the two major limitations of the ZI-DCT0 which are the randomness in choosing the threshold and the type of trading (CT or TF). To overcome such limitation, the central idea behind the DCT1 is that the trader will learn from the historical dataset of an asset's price time series earlier to the process of choosing a threshold and a type of trading.

Algorithm 2 illustrates the core trading mechanism for a DCT1 trader. Prior to trading in the market, a DCT1 trader will examine the status of the asset's price movements using the historical price dataset for the asset. Such examination aims for defining to great extend the most fitted practical threshold and type of trading for the trader as regard to the profitability.

In detail, a DCT1 trader will examine the profitability of the asset in terms of its historical price data using a defined number of verity thresholds which are generated randomly within a defined range. For each threshold value, the DCT1 trader will examine the historical price data set for the asset using the directional-change event approach from two points of view firstly as a CT while and secondly as a TF trader (as described in Section III). Subsequent to the DCT1 trader making a trade, the Rate of Investment (ROI) as a performance indicator will be computed. ROI is a performance measure defined as the total return to a trader's investment over a defined period, divided by the cost of the investment. The ROI is expressed as a percentage, and is either positive or negative,

 ISBN: 978-1-61208-390-2

Algorithm 2 The core trading mechanism for the DCT1.

Require: initialise variables ($e = upturnEvent$, $x = p_0$, $highestROI = 0$)
Input (p_n, λ_{min}, λ_{max}) // p_n training price dataset is used to train trader to find the best investment threshold and type of trading; n length of training dataset;$\lambda_{min}$is the minmum thrshold; and λ_{max} is the maximum thrshold.
For (i $= 0; i \prec 30; i++$) **do**// Examining 30 randomly generated thresholds
Begin
$\triangle x_{DC} = GenerateRandomThreshold[\lambda_{min}, \lambda_{max}]$
For ($y = 0; y \prec 2; y++$) **do** // Examining two trading types where $y = 0$ is CT and $y = 1$ is TF
Begin
For ($t = 0; t \prec n; t++$) **do** //Loop training price dataset
Begin
if ($e = upturnEvent$) **then**
if $p_t \leq x \times (1 - \triangle x_{DC})$**then**
$e \leftarrow downturnEvent$
$x \leftarrow p_t$
CT→Buy, TF→Sell
else
$x \leftarrow max(x, p_t)$
end if - Upturn event price examination
else //$e = downturnEvent$
if $p_t \geq x \times (1 + \triangle x_{DC})$ **then**
$e \leftarrow upturnEvent$
$x \leftarrow p_t$
CT→Sell, TF→Buy
else
$x \leftarrow min(x, p_t)$
end if - Downturn event price examination
endif - Event Examination
$ROI = Evaluate()$//ROI (rate of investment) is the result of evaluting the trader profit/loss for the given values of $\triangle x_{DC}$and y.
end for - End loop training price dataset
if ($ROI \succ highestROI$)**then**
$\lambda = \triangle x_{DC}$ // best threshold λ
$\omega = y$ // best type of trading ω
$highestROI = ROI$
endif
end for - End loop trading type
end for - End loop random threshold

which means that correspondingly, the trader achieves either a profit or makes a loss.

Towards the end of the examination, the threshold and the type of trading that results in the most profitable outcome with reference to the ROI will be chosen for the DCT1 trader's decision with regard to placing an order.

V. Experiments Description

In this section, we report on the experiments undertaken in the Agent-Based FX Market (ABFXM) that we developed. Our aim in particular, is to examine the profitability of the two strategies in term of the agents' return of investments; this can subsequently inform the design of trading strategies and decision support systems for the trading in the financial market.

A. Dataset

In this study, we used a high-frequency dataset (HFD) for EUR/USD historical prices provided by OANDA Corporation which is an online foreign currencies trading platform. HFD in finance refers to an extremely huge quantity of data which is the complete record of transactions and their associated characteristics at frequencies higher than on a daily basis [53]. According to Dacarogna et al. "The number of observations in one single day of a liquid market is equivalent to the number of daily data within 30 years" ([9], p. 6).

The dataset contains data samples of EUR/USD prices spanning the year of 2008 where each record contains three fields: (a) a bid and (b) an ask EUR/USD price at (c) a timestamp. This dataset is fed into the ABFXM via the market-maker. The time-span of the price dataset is very important in the study, given that different amounts of data examination possibly will provide ratios of precision interesting to study.

B. Agent-Based FX Market

In this section, we provide an overview of the ABFXM [15], which was developed to simulate the intraday trading activity at the level of an FX market-maker market. For a further detailed description of the ABFXM design, we refer interested reader to [15].

The FX market is where the exchange of currencies in which buying and selling currencies takes place. It is a decentralized market and operates 24 hours a day hence it is considered the largest and most liquid financial market in the world. The FX market is not an individual market given that it is composed of a global network of market-maker markets that connect investors from all around the world. Investors can be governments, central and commercial banks, institutional and individual investors, etc. Generally FX trading firms are market-makers [9]. A market-maker is a firm which supplies liquidity for currencies, and subsequently quotes both a buy and a sell price for a currency on its platform. The market-maker buys from and sells to its investors as well as other market-makers accordingly makes earnings from the difference between the bid and the offer price.

The ABFXM developed in [15] populated with N trading agents who participate in the market by means of buying and selling currencies. For simplicity, there is one currency pair (EUR/USD) available for trading in the ABFXM. A currency pair in the ABFXM is represented as base/quote wherein these two currencies are traded. We denote b_t at time t as the bid price by which a trading agent j can sell the base currency and buy an equivalent amount of the quoted currency to buy the base currency. This means agent j is opening a short position. Similarly, the ask price a_t at time t is the offer price at which agent j can buy the base currency and sell an equivalent amount of the quoted currency to pay for the base currency, opening a long position. During the market run, the market-maker uses a historical high-frequency EUR/USD

 ISBN: 978-1-61208-390-2

prices dataset to issue price quotes and feed these prices into the market. The prices are fed over a defined one month period and therefore the trading agents act in response.

Each trading agent is capable of holding, at time t during the market run, two different types of asset: a risk free asset (cash), and a risky asset (currency). Before the ABFXM launch, based on a continuous uniform distribution, each individual trading agent will be assigned a home currency and a margin ratio. Every trading agent j has a portfolio expressed in its home currency. The portfolio records the results of the agent's transactions during the market run. The Net Asset Value (NAV) at time t denoted by $NAV_{j,t}$ represents the current cash value of an agent j's account. In particular, the $NAV_{j,t}$ is the amount of cash in the agent j's account plus all unrealized profits and minus all unrealized losses associated with all the account's open positions.

The clearance mechanism of the ABFXM is simple where every market order at time t will be totally executed, whereas limit orders will be executed when their constraints are satisfied. A market order is an order for immediate execution in the market at the current price of the currency. On the contrary, a limit order is an order in which an agent a specifies the price at which it is willing to buy/sell a number of currency. An update will take place for each agent's portfolio that has an executable order at time t . Afterward, the market's time turns from t to $t+1$. Thus, the bid and ask prices are adjusted to the prices at time $t+1$ using the historical bid and ask prices. Hence based on the recent bid and ask prices, the portfolio will be updated for each agent holds an open position at time $t+1$. Finally, each open position at time $t+1$ will be verify for a margin call, which is a procedure to close out the agent's open position once the amount of cash in its account is under the minimum margin required to cover the size of its currently open position. The purpose of the margin call act is to stop an agent from losing more than the amount of cash available in its account.

C. Assumption

We make the following six assumptions in modelling the agents' trading mechanism:

Assumption 1 We assume that the trading agents endowed with 10,000 amounts of cash and without any shares.

Assumption 2 We assume that a position cannot be adjusted.

Assumption 3 A position is only opened by a market order or a limit order.

Assumption 4 We restrict the quantity of positions held by an agent a at time t to be one opened position.

Assumption 5 The market does not imply fees for the transactions.

Assumption 6 A trading agent invests 100% of it cash when buying, and 100% of it shares when selling.

In essence, the most important reason for these simplification assumptions is that by making these assumptions the complexity of the trading strategy is reduced to a level that can be studied and compared within the scope of this work. Simplicity and unification the initial variable of the agents' characteristics is fundamental block to clearly compare the efficiency of the two trading strategies. Therefore, allocating variable quantities results in a substantial complication of the comparison analysis. The relaxation of these six assumptions does not affect the generality of the simulation results shown in our paper. However, we are aware of the importance of the role of quantity and diversity as a choice variable.

Table I
ROI RESULTS FROM THE SIMULATION

Trading Strategy	Trend Type	AVG. Threshold	ROI
ZI-DCT0	TF	0.7 %	0.5 %
ZI-DCT0	CT	0.7 %	- 8.9 %
DCT1	TF	0.9 %	6.2 %

D. Results

One way to evaluate the performance of the DCT1 trading strategy is to look at the trading profits it generates. Therefore, we used the ROI as a performance indicator. The simulation uses historical bid and ask prices for the EUR/USD currency pair over a defined six month period during 2008, by feeding these prices into the market via the market-maker, and having the agents act in response to the price changes. The learning process for the DCT1 traders is over a four month period. The results generated from the simulation run are averaged over 10 independent simulation runs, each run adopting different initial seeds provided by random number generators, and different ranges of threshold values. We performed each independent simulation run with the same parameter configuration values, but with different seeds and ranges of threshold values, to ensure that the results of the simulation are consistent; this allows us to establish the robustness and accuracy of the simulation results.

We report the ROI results from the simulation run for three investment strategies: (i) TF ZI-DCT0; (ii) CT ZI-DCT0 and (iii) DCT1. The comparison of the performance of DCT1 with that of the ZI-DCT0 over a six month period is given in Table I. For the sample test period, the average ROI of the DCT1 is 6.2%, while for ZI-DCT0 TF is only 0.5%. An important observation to highlight is that the detected DCT1 threshold value and type of trading for the different simulation runs are roughly the same. Thus, this confirms that financial price time series exhibit periodic patterns. This is a good starting result regarding automats' trading strategies; though a full study through comparison with different trading strategies over different time periods and using different assets price time series will be more comprehensive, as this will show the full picture and effectiveness of the adopted trading strategy.

VI. Conclusion and Future Work

In this paper, we proposed a new trading strategy (DCT1) designed as a decision making support system tool for financial investors. It is derived from the ZI-DCT0. The main contribution of this paper is the combination of classical TF and CT investment rules and a learning model from historical prices with regard to financial time series, which can significantly

 ISBN: 978-1-61208-390-2

improve computational effectiveness and the predictability of price trend directions, and uncover periodic patterns. TF and CT strategies, owing to their investment efficiency, have been widely adopted by investors. To the best of our knowledge, no related research in the literature has investigated TF and CF investment strategies within a learning model based on the detection of periodic directional change patterns. This study has demonstrated the feasibility of employing learning as part of a trading strategy in that DCT1 is designed to adapt to market price trend directions and hence deduct periodic patterns. Future work will consider the combination of evolutionary learning techniques with a DC event approach for developing trading strategies for investment in financial markets.

ACKNOWLEDGMENTS

We would like to thank the Deanship of Scientific Research in King Saud University for their support. We would like to thank the OANDA Corporation for providing the FX market datasets. We are very grateful to Prof. M. Fasli, Prof. E. Tsang, Prof. R. Olsen and Dr. A. Dupuis for their insightful suggestions and advice during the course of this study. We would also like to thank the anonymous reviewers for their useful comments and suggestions.

REFERENCES

[1] S. Leroy, "Risk aversion and the martingale property of stock prices," *International Economic Review*, vol. 14 (2), pp. 436–446, 1973.

[2] A. Beja, "The limits of price information in market processes," Tech. Rep. 61, Research Program in Finance, University of California, Berkeley, 1977.

[3] R. Lucas, "Asset prices in an exchange economy," *Econometrica*, vol. 46(6), pp. 1429–1445, 1978.

[4] S. Grossman and J. Stiglitz, "On the impossibility of informationally efficient markets," *The American Economic Review*, vol. 70, no. 3, pp. 393–408, 1980.

[5] J. Tirole, "On the possibility of speculation under rational expectations," *Econometrica*, vol. 50 (5), pp. 1163–1181, 1982.

[6] A. Lo, "Stock market prices do not follow random walks: evidence from a simple specification test," *Review of Financial Studies*, vol. 1 (1), pp. 41–66, 1988.

[7] E. Tsang and S. Martinez-Jaramillo, "Computational finance," *IEEE Computational Intelligence Society Newsletter*, pp. 3–8, 2004.

[8] R. Cont, "Empirical properties of asset returns: stylized facts and statistical issues," *Quantitative Finance*, vol. 1, no. 2, pp. 223–236, 2001.

[9] M. Dacorogna, R. Gençay, U. Müller, R. Olsen, and O. Pictet, *An introduction to high-frequency finance*. San Diego: Academic Press, 2001.

[10] M. Covel, *Trend following: How great traders make millions in up or down markets*. Financial Times Prentice Hall, 2004.

[11] S. Fong, J. Tai, and Y. W. Si, "Trend following algorithms for technical trading in stock market," *Journal of Emerging Technologies in Web Intelligence*, vol. 2, no. 3, pp. 136–145, 2011.

[12] S. Fong, Y. Si, and J. Tai, "Trend following algorithms in automated derivatives market trading," *Expert Systems with Applications*, vol. 39, no. 13, pp. 11378–11390, 2012.

[13] A. Szakmarya, Q. Shenb, and S. Sharmac, "Trend-following trading strategies in commodity futures: A re-examination," *Journal of Banking & Finance*, vol. 34, no. 2, pp. 409–426, 2010.

[14] J. James, "Simple trend-following strategies in currency trading," *Quantitative Finance*, vol. 3, pp. 75–77, 2003.

[15] M. Aloud, E. Tsang, and R. Olsen, "Modelling the FX market traders' behaviour: an agent-based approach," in *Simulation in Computational Finance and Economics: Tools and Emerging Applications* (B. Alexandrova-Kabadjova, S. Martinez-Jaramillo, A. Garcia-Almanza, and E. Tsang, eds.), Hershey, Pennsylvania: IGI Global, 2012.

[16] M. Aloud, E. Tsang, R. Olsen, and A. Dupuis, "A directional-change events approach for studying financial time series," *Economics Papers*, vol. No 2011-28, 2011.

[17] V. Alfi, M. Cristelli, L. Pietronero, and A. Zaccaria, "Minimal agent based model for financial markets I: origin and self-organization of stylized fatcs," *The European Physical Journal B*, vol. 67, no. 3, pp. 385–397, 2009.

[18] G. Daniel, *Asynchronous simulations of a limit order book*. PhD thesis, University of Manchester, 2006.

[19] J. Duffy and M. Unver, "Asset price bubbles and crashes with near-zero-intelligence traders," *Economic Theory*, vol. 27, pp. 537–563, 2006.

[20] D. Gode and S. Sunder, "Allocative efficiency of markets with zero intelligence (Z1) traders: market as a partial substitute for individual rationality," *Journal of Political Economy*, vol. 101, no. 1, pp. 119–137, 1993.

[21] S. Martinez-Jaramillo and E. Tsang, "An heterogeneous, endogenous and co-evolutionary GP-based financial market," *IEEE Transactions on Evolutionary Computation*, vol. 13, no. 1, pp. 33–55, 2009.

[22] D. Cliff and J. Bruten, "More than zero intelligence needed for continuous double-auction trading," Tech. Rep. HPL-97-157, HP Laboratories Bristol, 1997.

[23] D. Gode and S. Sunder, "Lower bounds for efficiency of surplus extraction in double auctions," in *The Double Auction Market: Institutions, Theories, and Evidence. Santa Fe Institute Studies in the Sciences of Complexity* (D. Friedman and J. Rust, eds.), pp. 199–219, Cambridge: Perseus Publishing, 1993.

[24] V. Alfi, M. Cristelli, L. Pietronero, and A. Zaccaria, "Minimal agent based model for financial markets II: statistical properties of the linear and multiplicative dynamics," *The European Physical Journal B*, vol. 67, no. 3, pp. 399–417, 2009.

[25] V. Alfi, M. Cristelli, L. Pietronero, and A. Zaccaria, "Mechanisms of self-organization and finite size effects in a minimal agent based model," *J. Stat. Mech*, vol. P03016, 2009.

[26] B. Wong and Y. Selvi, "Neural network applications in finace: a review and analysis of literature," *Information & Management*, vol. 34, pp. 129–139, 1998.

[27] J. Holland, *Adaptation in Natural and Artificial Systems*. Ann Arbor, MI: University of Michigan Press, 1975.

[28] J. Koza, *Genetic Programming: on the Programming of Computers by Means of Natural Selection*. Cambridge: The MIT Press, 1992.

[29] S. Schulenburg, P. Ross, and S. Bridge, "Strength and money: an LCS approach to increasing returns," in *Advances in Learning Classifier Systems*, vol. 1996 of *Lecture Notes in Artificial Intelligence*, pp. 114–137, Springer-Verlag, 2000.

[30] S. Schulenburg and P. Ross, "Explorations in LCS models of stock trading," in *Advances in Learning Classifier Systems, Lecture Notes in Artificial Intelligence*, no. 2321, pp. 150–179, Springer-Verlag, 2002.

[31] M. Austin, G. Bates, M. Dempster, V. Leemans, and S. Williams, "Adaptive systems for foreign exchange trading," *Quantitative Finance*, vol. 4, no. 4, pp. 37–45, 2004.

[32] E. Azoff, *Neural Network Time Series Forecasting of Financial Markets*. New York, NY, USA: John Wiley & Sons, Inc., 1994.

[33] A. Refenes, *Neural Networks in the Capital Markets*. New York, NY, USA: John Wiley & Sons, Inc, 1994.

[34] R. Trippi and E. Turban, *Neural networks in finance and investing: using artificial intelligence to improve realworld performance*. Burr Ridge: Irwin Professional Publishing Co., 1996.

[35] G. Zhang, B. Patuwo, and M. Hu, "Forecasting with artificial neural networks: the state of the art," *International Journal of Forecasting*, vol. 14, pp. 35–62, 1998.

[36] Y. Shachmurove, "Business applications of emulative neural networks," *International Journal Of Business*, vol. 10, 2005.

[37] J. Yao and C. Tan, "A case study on using neural networks to perform technical forecasting of forex," *Neurocomputing*, vol. 34, no. 1-4, pp. 79–98, 2000.

[38] S. Hayward, "Genetically optimized artificial neural network for financial time series data mining," in *Simulated Evolution and Learning* (T.-D. Wang, X. Li, S.-H. Chen, X. Wang, H. Abbass, H. Iba, G.-L. Chen, and X. Yao, eds.), vol. 4247 of *Lecture Notes in Computer Science*, pp. 703–717, Springer Berlin / Heidelberg, 2006.

[39] R. Batchelor and G. Albanis, "Predicting high performance stocks using dimensionality reduction techniques based on neural networks," in *Developments in Forecast Combination and Portfolio Choice* (C. Dunis and A. Timmerman, eds.), pp. 117–134, Kluwer Academic Publishers, 2001.

 ISBN: 978-1-61208-390-2

[40] M. Dempster, T. Payne, Y. Romahi, and G. Thompson, "Computational learning techniques for intraday fx trading using popular technical indicators," *IEEE Transactions on Neural Networks*, vol. 12, no. 4, pp. 744–754, 2001.

[41] S. Gutjahr, M. Riedmiller, and J. Klingemann, "Daily prediction of the foreign exchange rate between the US dollar and the German mark using neural networks," in *Joint PACES / SPICES Conference*, pp. 492–498, 1997.

[42] J. Yao, C. Tan, and Y. Li, "Option prices forecasting using neural networks," *Omega: The International Journal of Management Science*, vol. 28, pp. 455–466, 2000.

[43] H. Zimmermann, R. Neuneier, and R. Grothmann, "Multi-agent modeling of multiple FX-markets by neural networks," *IEEE Transactions on Neural Networks, Special issue*, vol. 12, no. 4, pp. 735–743, 2001.

[44] F. Allen and R. Karjalainen, "Using genetic algorithms to find technical trading rules," *Journal of Financial Economics*, vol. 51, pp. 245–271, 1999.

[45] R. Bauer, *Genetic Algorithms and Investment Strategies*. New York: John Wiley & Sons, 1994.

[46] W. Leigh, R. Purvis, and J. Ragusa, "Forecasting the NYSE composite index with technical analysis, pattern recognizer, neural networks, and genetic algorithm: a case study in romantic decision support," *Decision Support Systems*, vol. 32, pp. 361–377, 2002.

[47] S. Mani, "Financial forecasting using genetic algorithms," *Applied Artificial Intelligence*, vol. 10, pp. 543–566, 1996.

[48] T. Lux, , and S. Schornstein, "Genetic learning as an explanation of stylized facts of foreign exchange markets," *Journal of Mathematical Economics*, vol. 41, no. 1-2, pp. 169–196, 2005.

[49] M. Versace, R. Bhatt, O. Hinds, and M. Shiffer, "Predicting the exchange traded fund dia with a combination of genetic algorithms and neural networks," *Expert Systems with Applications*, vol. 27, pp. 417–425, 2004.

[50] J. Holland and J. Miller, "Artificial adaptive agents in economic theory," *The American Economic Review*, vol. 81, pp. 365–370, 1991.

[51] W. B. Arthur, J. H. Holland, B. LeBaron, R. Palmer, and P. Tayler, "Asset pricing under endogenous expectations in an artificial stock market," in *The economy as an evolving, complex system II* (W. Arthur, D. Lane, and S. Durlauf, eds.), pp. 15–44, Redwood City, CA: Addison Wesley, 1997.

[52] M. Aloud and M. Fasli, "The impact of strategies on the stylized facts in the fx market," tech. rep., University of Essex, United Kingdom, 2013.

[53] R. Engle, "The econometrics of ultra-high frequency data," *Econometrica*, vol. 68, no. 1, pp. 1–22, 2000.

 ISBN: 978-1-61208-390-2

Leader-Following Formation Control with an Adaptive Linear and Terminal Sliding Mode Combined Controller Using Auto-Structuring Fuzzy Neural Network

Masanao Obayashi, Kohei Ishikawa,
Takashi Kuremoto, Shingo Mabu
Graduate School of Science and Engineering,
Yamaguchi University,Ube, Yamaguchi, Japan
email:{m.obayas, s003vk, wu, mabu}@yamaguchi-u.ac.jp

Kunikazu Kobayashi
School of Information Science and Technology,
Aichi Prefectural University, Nagakute, Aichi, Japan
email: kobayashi@ist.aichi-pu.ac.jp

***Abstract*—This paper proposes an intelligent formation control method of the leader and follower agents with nonlinear dynamics. In the proposed method, agents can exchange only information of positions of available agents and the follower agents follow the leader, taking a predefined formation. In the real world, the method that doesn't need a lot of information to make cooperative behaviors is very useful for the case of environments existing communication delay and weak communication. In addition, each agent will be controlled by the linear and terminal combined sliding mode control method. Furthermore, to adapt the change of the environment, the Auto-Structuring Fuzzy Neural Control System (ASFNCS) is introduced to provide appropriate control inputs while coping well with disturbance and nonlinear dynamics. In the simulation, it is verified that the proposed method is useful in the point of performance of the leader-following formation control.**

Keywords-formation control; linear sliding mode control; terminal sliding mode control; auto-structuring fuzzy neural network

I. INTRODUCTION

Recently, formation problems have attracted more researchers as their importance and necessity are known widely. For example, Hou et al. [1] proposed a robust adaptive control approach to solve a consensus problem of Multi Agent Systems (MAS), Song et al. [2] and Cheng et al. [3] dealt with leader-follower consensus problem. Defoort et al. [4], Morbidi et al. [5] proposed a formation following control using sliding mode control, a kind of representative robust control methods. Cui et al, [6] dealt with leader-follower formation control of underactuated autonomous underwater vehicles. Yu et al. [7] dealt with time-varying velocity cases of a distributed leader-follower flocking control for multi-agent dynamical systems. In the view point of controllers using soft computing tools, Chang et al. [8] used fuzzy, Lin [9] used fuzzy basis function network, Chen et al. [10] used neural networks, Lin et al. [11] used reinforcement Q learning method, for formation control. However, designed controller structures for these formation or consensus problems are fixed, therefore, their methods are no flexible for changes of environments and they have possibility of useless computation load. Many of the control methods mentioned above are using linear sliding mode control method. Recently, terminal sliding mode control method has been used for constructing the following controller because of its superiority on fast finite time convergence and less steady state errors to objective states of the agents, Yu et al. [12], Zou et al. [13], Chang et al. [14].

We have already proposed the leader-following formation control method, Obayashi et al. [15] that had the following three features; use of linear sliding mode control, and the controller had the optimal structure, named ASFNCS, using self-structuring algorithm adapting to change of environments making use of the concept of Cheng et al. [16] and then introducing of making only use of the positions of agents without velocities of them.

In this paper, we consider to apply the superior point of the terminal sliding mode control to our previous work, Obayashi et al. [15], that is, to switch the linear sliding mode control and terminal one appropriately.

The rest of the paper is as follows: in Section II preparation to be required to follow the paper. The structure of the proposed system are described in Section III. The controller design is described in Section IV. In Sections V, VI, and VII, learning algorithm, following control problem and computer simulation are described, respectively. We have a conclusion in Section VIII.

II. PREPARATION

The dynamics of each agent consisting of MAS is described as following the $n\,th$ order nonlinear differential equation,

$$\begin{aligned} x^{(n)} &= f(\mathbf{x}, u_{ASFNS}) + \Delta f(\mathbf{x}, u_{ASFNS}) + d \\ &= f(\mathbf{x}, u_{ASFNS}) - h \cdot u_{ASFNS} + h \cdot u_{ASFNS} + \delta_1 , \\ &= f_n(\mathbf{x}, u_{ASFNS}) + h \cdot u_{ASFNS} + \delta_1 \end{aligned} \quad (1)$$

where $\mathbf{x} = [x \quad \dot{x} \quad \cdots \quad x^{(n-1)}]^T \in R^n$: observable states of the agent, $f(\mathbf{x})$: unknown continuous function, h : a predefined constant, u_{ASFNS} : control input, δ_1 : term including the indefinite term $\Delta f(\mathbf{x}, u_{ASFNS})$ of the agent and the disturbance. δ_1 is assumed that $|\delta_1| < D_1$. Here, D_1 is a predefined positive constant.

 ISBN: 978-1-61208-390-2

III. STRUCTURE OF THE PROPOSED SYSTEM

The proposed system in this paper is shown in Figure 1.

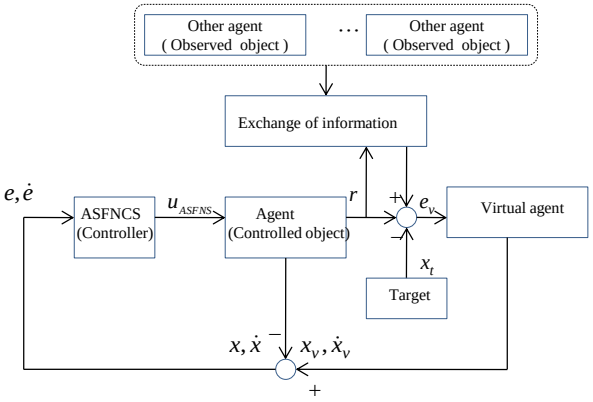


Figure 1. Structure of the proposed multi agent control system.

The objective of the control: The derivation of the control law which makes the orbit $\mathbf{x}$ of the states of the agent follow the reference orbit $\mathbf{x}_v$ accurately. We define the following error vector $\mathbf{e} = [e, \dot{e}, \cdots, e^{(n-1)}]$ as

$$\mathbf{e} = \mathbf{x}_v - \mathbf{x}, \tag{2}$$

where $\mathbf{x}_v$: the reference vector, that is, their virtual agent state vector followed by its own agent. The virtual agent has its position and velocity decided by observing the position of the leader (see Section VI).

IV. CONTROLLER DESIGN

Figure 2 shows the ASFNCS. The ASFNCS consists of the fuzzy neural network controller with function of node adding/pruning (Auto-structuring FNNC) and the robust controller.

A *Fuzzy neural network controller (FNNC)*

Figure 3 shows the structure of the FNNC. Γ_k , output of k th node of the hidden layer, means the fitness of rule k .

$$\Gamma_k(\mathbf{e}) = \prod_{i=1}^{n} \exp\left\{ -\frac{\left(e^{(i-1)} - m_{ki}\right)^2}{\left(\sigma_{ki}\right)^2} \right\} \quad , \tag{3}$$

$$(k = 1, \cdots, R) \ ,$$

$$u_{asfnn} = \sum_{k=1}^{R} \xi_k \cdot \Gamma_k , \tag{4}$$

where $\mathbf{m}_k = [m_{ki}, m_{k2}, \cdots, m_{kn}]$, $\boldsymbol{\sigma}_k = [\sigma_{ki}, \sigma_{k2}, \cdots, \sigma_{kn}]$ are center and width vectors of the kth rule output function, respectively. R is the number of the rule nodes. u_{asfnn} is the output of the FNNC. ξ_k is the weight between the k *th* rule node and the output of the FNNC.

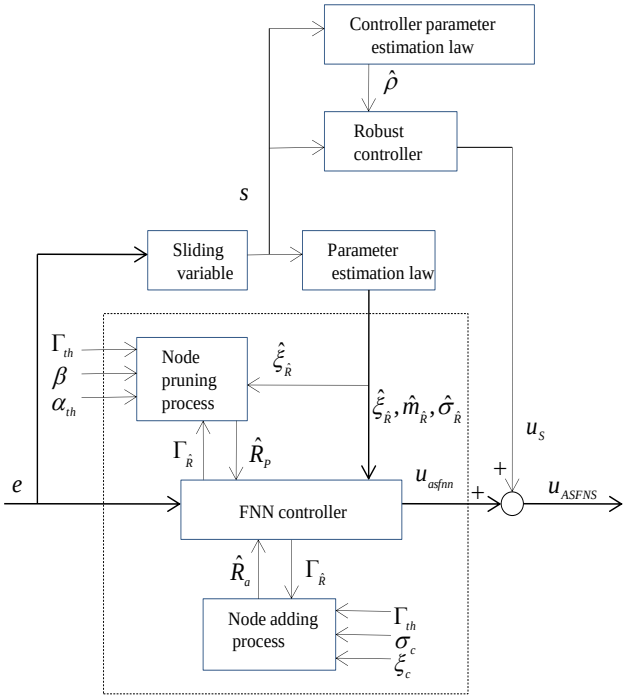


Figure 2. Structure of theASFNCS.

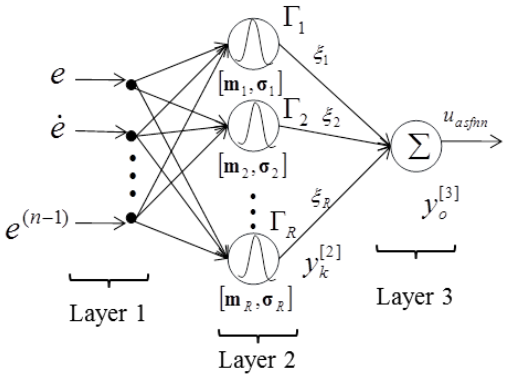


Figure 3. Structure of the FNNC.

The output u^* of ideal controller is like that,

$$u^* = u^*_{asfnn} + \varphi_1 = \sum_{k=1}^{R^*} \xi_k^* \Gamma_k + \varphi_1, \tag{5}$$

φ_1: difference between ideal controller and aproximated optimal FNNC.

However, u^* would not be used for controller design actually (see Sectiuon IV *B*).

The estimated value $\hat{u}_{asfnn}$ of the optimal FNNC u^*_{asfnn} is as follows,

$$\hat{u}_{asfnn} = \sum_{k=1}^{\hat{R}} \hat{\xi}_k \cdot \Gamma_k , \tag{6}$$

where $\hat{\xi}_k$ is the estimate value of the optimal ξ_k^*.

B. Auto structuring mechanism of FNNC

Node adding: Nodes adding process is as follows,

$$\Gamma_{\max} = \max(\Gamma_k), \quad k = 1,2,\cdots,R(t), \tag{7}$$

where $R(t)$ is the number of the node at time t. When the next relation exits,

$$\Gamma_{\max} \le \Gamma_{th}, \tag{8}$$

where $\Gamma_{th} \in [0,1]$ is a pre-defined threshold and given by trial and error, a new node σ_c is added and the center and width of the node are like this,

$$m_{R_a i} = e^{(i-1)}, \sigma_{R_a i} = \sigma_c, \ (i = 1,\cdots,n), \tag{9}$$

where σ_c is a predefined positive constant and the weight ξ_{R_a} between the new node and the output is set as

$$\xi_{R_a} = \xi_c, \tag{10}$$

where R_a is number of the added node.

Node pruning: Nodes pruning process is as follows, the cost function for measuring the important index α_k of the node k is defined as,

$$E_1 = \frac{1}{2}(u^* - \sum_{k=1}^{\hat{R}} \hat{\xi}_k \Gamma_k)^2, \tag{11}$$

using the Taylor series expansion, where the parameter falls intou the local minimum, the sensitivity of E_1 for $\dot{\hat{\xi}}_k$ becomes next equation,

$$E_1(\dot{\hat{\xi}}_k) \approx \frac{1}{2}(\dot{\hat{\xi}}_k \Gamma_k)^2, \quad k = 1,2,\ldots,\hat{R}. \tag{12}$$

Here, we define the important index of the node as

$$\alpha_k(t+1) = \alpha_k(t) \times \exp\left\{-\beta \times \delta[E_{th} - E_1(\dot{\hat{\xi}}_k)]\right\}, \tag{13}$$

and $\delta[*]$ function to decide whether make α_k decrease is defined as

$$\delta[E_{th} - E_1(\dot{\hat{\xi}}_k)] = \begin{cases} 1 & \text{if } E_{th} - E_1(\dot{\hat{\xi}}_k) \ge 0 \\ 0 & \text{if } E_{th} - E_1(\dot{\hat{\xi}}_k) < 0 \end{cases}. \tag{14}$$

This means that if the rule k is not activated enough, that is, $E_{th} - E_1(\dot{\hat{\xi}}_k) \ge 0$, then, α_k decreases. If $\alpha_k \le \alpha_{th}$, the node k is considered as unnecessary node and it is deleted. Here α_{th} is a predefined positive constant and given by trial and error. The number of deleted nodes is presented as $\hat{R}_p$. The output u_{asfnn} of the FNNC introducing the auto-structuring mechanism is as follows,

$$u_{asfnn} = \underbrace{\sum_{k=1}^{\hat{R}} \hat{\xi}_k \Gamma_k}_{u_R} - \underbrace{\sum_{p=1}^{\hat{R}_p} \hat{\xi}_p \Gamma_p}_{u_p}, \tag{15}$$

where u_R is the output after adding new nodes and, u_p is the output after pruning nodes. With (15), we get

$$\tilde{u}_{asfnn} = u^* - \hat{u}_{asfnn} = \sum_{k=1}^{\hat{R}} \tilde{\xi}_k \Gamma_k + \varepsilon, \tag{16}$$

where $\tilde{\xi} = \xi^* - \hat{\xi}$.

C. Stability analysis

The structure of the auto-structuring fuzzy neural control system (ASFNCS) is shown in Figure 2. The output u_{ASFNS} of the ASFNCS is as follows, Cheng et al. [16],

$$u_{ASFNS} = u_{asfnn} + u_s, \tag{17}$$

u_{asfnn} is described by (15), and u_s is the output of the robust controller. To derive an adaptive controller, we define the sliding variable s as

$$s = e^{(n-1)} + k_1 e^{(n-2)} + \cdots + k_{n-1} e + \gamma\, k_{n1} e^{p/q} + (1-\gamma) k_{n2} \int_0^t e dt, \tag{18}$$

where each of all the cofficients k_* is pre-defined positive constant, $\gamma (0 \le \gamma \le 1)$ is a parameter. When $\gamma = 0.0,\ 1.0$, sliding variable (18) can be regarded as LSM, TSM, respectively. In this paper, γ changes nonlinearly from 1.0 to 0.0 during the controller working.

The p and q are positive odd integers, which satisfy the following condition:

$$p > q. \tag{19}$$

Then, a sliding mode controller can be designed as follows :

$$u_{SMC} = u_E + u_H, \tag{20}$$

An equivalent controller u_E is expressed as

$$u_E = h^{-1}(-f_{n(x,u)} + x_v^{(n)} + k_1 e^{(n-1)} + \cdots + k_{n-1}\dot{e} + \gamma\, k_{n1} \frac{q}{p} e^{q/p} \cdot \dot{e} + (1-\gamma) k_{n2} \cdot e) \tag{21}$$

$$u_H = D_1 \operatorname{sgn}(s), \tag{22}$$

Using (1), (2),(18) and (21), we get

$$\dot{s} = e^{(n)} + + k_1 e^{(n-1)} + \cdots + k_{n-1}\dot{e} + \gamma\, k_{n1} \frac{q}{p} e^{q/p} \cdot \dot{e} + (1-\gamma) k_{n2} \cdot e = h(u^* - u_{ASFNS}). \tag{23}$$

Using Eq. (16), Eq. (23) can be rewritten as

$$\dot{s} = h(u^* - u_{asfnn} - u_s) = h(\sum_{k=1}^{\hat{R}} \tilde{\xi}_k \Gamma_k + \varepsilon - u_s). \quad (24)$$

In this paper, the robust controller is used to eliminate the effect of error ε. Consider the Lyapunov function candidate in the following form:

$$V = \frac{1}{2}s^2 + \frac{h}{2\eta_\xi}\sum_{k=1}^{\hat{R}}\tilde{\xi}_k^2 + \frac{h}{2\eta_\rho}\tilde{\rho}^2. \quad (25)$$

Take the derivative of Eq.(25) and using Eq.(24), it is concluded that

$$\begin{aligned}\dot{V} &= s\dot{s} + \frac{h}{\eta_\xi}\sum_{k=1}^{\hat{R}}\tilde{\xi}_k\dot{\tilde{\xi}}_k + \frac{h}{\eta_\rho}\tilde{\rho}\dot{\tilde{\rho}} \\ &= h\left\{\sum_{k=1}^{\hat{R}}\tilde{\xi}_k\left(s\Gamma_k + \frac{\dot{\tilde{\xi}}}{\eta_\xi}\right) + s(\varepsilon - u_s) + \frac{1}{\eta_\rho}\tilde{\rho}\dot{\tilde{\rho}}\right\}.\end{aligned} \quad (26)$$

For achieving $\dot{V} \le 0$, the adaptation law and the robust controller are chosen as

$$\dot{\tilde{\xi}} = -\dot{\hat{\xi}}_k = -\eta_\xi s\Gamma_k, \quad k = 1,2,\ldots,\hat{R}, \quad (27)$$

$$u_s = \hat{\rho}\,\mathrm{sgn}(s), \quad (28)$$

$$\dot{\tilde{\rho}} = -\dot{\hat{\rho}} = -\eta_\rho|\rho|, \quad (29)$$

where η_ρ and η_ξ are positive constants. Then Eq. (26) can be rewritten as

$$\begin{aligned}\dot{V} &= h\{s\varepsilon - \hat{\rho}|s| - (\rho - \hat{\rho})|s|\} \le h\{|\varepsilon||s| - \rho|s|\} \\ &= -h(\rho - |\varepsilon|)|s| \le 0\end{aligned}. \quad (30)$$

If $|\varepsilon| \le \rho$, $\dot{V} \le 0$ holds. By Barbalat's Lemma, $s \to 0$ as $t \to \infty$, then the stability is guaranteed.

V. LEARNING ALGORITHM

We introduce the online learning algorithm to adjust the parameters of ASFNS, Cheng et al. [16]. We derive the algorithm the gradient-decent method. First, the adaptive law shown in (27) can be written as

$$\dot{\hat{\xi}}_k = \eta_\xi s\Gamma_k. \quad (31)$$

According to the gradient-decent algorithm, the adaptive law of $\hat{\xi}$ also can be represented as

$$\dot{\hat{\xi}} = \eta_\xi\frac{\partial E_2}{\partial \hat{\xi}_k} = \eta_\xi\frac{\partial E_2}{\partial y_o^{[3]}}\frac{\partial y_o^{[3]}}{\partial net_o^{[3]}}\frac{\partial net_o^{[3]}}{\partial \hat{\xi}_k} = \eta_\xi\frac{\partial E_2}{\partial y_o^{[3]}}y_k^{[2]}, \quad (32)$$

where $E_2 = 1/2(x_v - x)^2$ is defined as the cost function. From Figure 3, (3), and (4), $net_k^{[2]} = \sum_{i=1}^{n}\left\{-(e^{(i-1)} - \hat{m}_{ki})^2/(\hat{\sigma}_{ki})^2\right\}$, $y_k^{[2]} = \exp(net_k^{[2]})$, $y_o^{[3]} = \sum_k \hat{\xi}_k\, y_k^{[2]} = \sum_k \hat{\xi}_k\,\Gamma_k$, $*$ in the [*] means the number of layer in Figure 3. Thus the Jacobian term $\partial E_2/\partial y_o^{[3]} = s$ is obtained through observation of (31) and (32). Thus, the adaptive law for the estimation terms of means $\hat{m}_{ki}$, and variance $\hat{\sigma}_{ki}$ can be derived as

$$\begin{aligned}\dot{\hat{m}}_{ki} &= \eta_m\frac{\partial E_2}{\partial \hat{m}_{ki}} = \eta_\xi\frac{\partial E_2}{\partial y_o^{[3]}}\frac{\partial y_o^{[3]}}{\partial y_k^{[2]}}\frac{\partial y_k^{[2]}}{\partial net_k^{[2]}}\frac{\partial net_k^{[2]}}{\partial \hat{\sigma}_{ki}} \\ &= \eta_m s\hat{\xi}_k\Gamma_k\frac{2(e^{(i-1)} - \hat{m}_{ki})^2}{(\hat{\sigma}_{ki})^2},\end{aligned} \quad (33)$$

$$\begin{aligned}\dot{\hat{\sigma}}_{ki} &= \eta_\sigma\frac{\partial E_2}{\partial \hat{m}_{ki}} = \eta_\xi\frac{\partial E_2}{\partial y_o^{[3]}}\frac{\partial y_o^{[3]}}{\partial y_k^{[2]}}\frac{\partial y_k^{[2]}}{\partial net_k^{[2]}}\frac{\partial net_k^{[2]}}{\partial \hat{\sigma}_{ki}} \\ &= \eta_m s\hat{\xi}_k\Gamma_k\frac{2(e^{(i-1)} - \hat{m}_{ki})^2}{(\sigma_{ki})^2},\end{aligned} \quad (34)$$

where η_m, η_σ are positive constants.

A. The algorithm of ASFNCS

Step 1 Initialize the parameters of ASFNCS.
Step 2 Calculate s using (18).
Step 3 Calculate u_{asfnn} using (15) and update $\hat{\xi}, \hat{m}, \hat{\sigma}$ using (32)-(34).
Step 4 Calculate u_s using (28) and adjust $\hat{\rho}$ using (29).
Step 5 Calculate u_{ASFNS} using (17) and input it to the agent.
Step 6 If the controlling time is over, the simulation ends, else go to Step 2.

VI. LEADER-FOLLOWING FORMATION CONTROL PROBLEM

The dynamics (1) of the $i\,th$ agent in multi agents is rewritten as

$$\ddot{x}_i(t) = f(t, x_i(t), \dot{x}_i(t)) + d_i + u_{i,ASFNC}, \quad (35)$$

where t : time, $x_i = (x_{1i}, x_{2i})^T$, $\dot{x}_i = (\dot{x}_{1i}, \dot{x}_{2i})^T$: the position, velocity of the agent, respectively. $f(t, x_i(t), \dot{x}_i(t)) \in R^2$: a term including the nonlinearity $u_{ASFNS} = (u_{1i}, u_{2i})^{\mathrm{T}} \in R^2$, $u_{i,ASFNC} = (u_{1i}, u_{2i})^T \in R^2$: control input to the $i\,th$ agent, $d_i \in R^2$: disturbance. In this paper, we consider the following control method that make the agents follow and catch the leader keeping the formation by the group, Cui et al. [6]. We assume that agents could exchange only positions information of the agents each other. Observing only the position of the leader, each agent constructs the virtual orbit to the leader. Realizing this virtual orbit by the virtual agent, and the agent follows its own agent. The following describes the method of following the leader by the virtual agent.

 ISBN: 978-1-61208-390-2

The following error of the $i\,th$ virtual agent is expressed as

$$e_{vi} = ((z_i + \omega_i) - x_t) + \sum_{j \in N_i} ((z_i + \omega_i) - z_j), \quad (36)$$

where e_{vi} :consensus error, $z_i = x_i$, $\omega_i = (x_{vi} - x_i)$, $x_t \in R^2$: the position of the leader in a 2-dimension space.

The following error x_{ei} of the $i\,th$ virtual agent considering the integral term to reduce the steady state error is expressed as

$$x_{ei} = e_{vi} + \phi_i + \int_0^t e_{vi} dt \ , \quad (37)$$

where $\phi_i = [\phi_{1i}, \phi_{2i}]^T$, the derivative of ϕ_i with respect to time is defined as

$$\dot{\phi}_i = -\beta_{1i}(\phi_i) - K x_{ei} \ , \quad (38)$$

where $\beta_{1i}(\phi_i) = [\lambda_1 \tanh(\phi_i / \lambda_1), \lambda_2 \tanh(\phi_i / \lambda_2)]^T$, $K = diag\,[\ k_{v1}, k_{v2}\ ]$, $\phi_i(0) = 0$.

The velocity $\dot{x}_{vi}$ of the $i\,th$ virtual agent is defined as

$$\dot{x}_{vi} = \beta_{1i}(\phi_i) + \beta_{2i}(\phi_i) \ , \quad (39)$$

where $\beta_{2i}(\phi_i) = [k_{v1} \tanh(\phi_i / k_{v1}), k_{v2} \tanh(\phi_i / k_{v2})]^T$, k_*, λ_* are positive constants.

VII. COMPUTER SIMULATION

We demonstrate the effectiveness of our proposed method by trying to following the leader (target) by the multi agent constructing by four agents.

The horizontal and vertical direction of velocity of the leader are $\dot{x}_{t1} = 0.02\cos(0.05t)$, $\dot{x}_{t2} = 0.015\sin(0.03t)$, respectively, and the initial position of the leader is given as $x_t = [0,0]^T$. The initial horizontal and vertical positions of the agents are both given randomly in the range of $[-5,5]$, and the initial velocities of all the four agents are equal to 0. The parameters used in the simulation are shown in Tables I and II. The dynamics and disturbances of the each agent are given as

$$f(t, x_i(t), \dot{x}_i(t)) = \begin{pmatrix} \sin(x_{1i}) + \dot{x}_{1i}\cos(\dot{x}_{2i}) \\ \cos(x_{2i}) + \dot{x}_{2i}\sin(\dot{x}_{1i}) \end{pmatrix} , \quad (40)$$

$$d_1 = \begin{pmatrix} -\sin(2t) \\ \sin(t^2) \end{pmatrix}, d_2 = \begin{pmatrix} \sin(t) \\ \cos(t) \end{pmatrix}, d_3 = \begin{pmatrix} \cos(2t) \\ -\cos(t) \end{pmatrix}, d_4 = \begin{pmatrix} -\sin(t) \\ -\cos(2t) \end{pmatrix} . \quad (41)$$

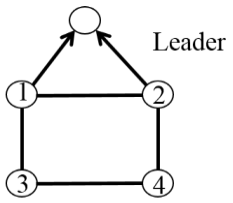


Figure 4. Network structure of multi agents.

The information exchanges among agents each other are carried out according to Figure 4. Each agent behaves dispersive and cooperatively through information exchanges and follows the leader, taking the predefined formation. In this simulation, from Figure 4, only agent 1 and 2 can observe the position of the leader and the agent 4 can observe the positions of the agent 1 and 3, and so on. Sampling time is $0.01[s]$ and total controlling time is 50 [sec].

Each of Figures 5-11 has 5 orbits; red is for target (leader), other 4 orbits are for follower agebts. The orbits of the target and follower agents in the case of the conventional method, LSM, are shown in Figure 5. Figure 6 is the enlarged figure of the transient (initial part of) positons of the leader and follower agents. The orbits of the leader and agents in the case of the conventional method, LSM, are shown in Figure 7. Figure 8 is the enlarged figure of the transient (initial part of) positons of the agents and the leader. Comparing these four figures, TSM method is superior to the LSM method in the point of consensus error of the positions of transient states of the leader and follower agents. However, the consensus error of them in the point of the steady state by LSM method is smaller than those of TSM method.

TABLE I. PARAMETERS OF THE VIRTUAL AGENT.

Parameter	Setting value
λ_1, λ_2	0.015
k_{v1}, k_{v2}	0.025

TABLE II. PARAMETERS OF THE ASFNCS.

Parameter	Setting value
k_1	2
k_2	3
Γ_{th}	0.3
σ_c	2
ξ_c	0
α_c	0.5
E_{th}	0.00001
β	0.002
α_{th}	0.05
η_ξ	1
η_ρ	0.02
η_m, η_σ	0.0015

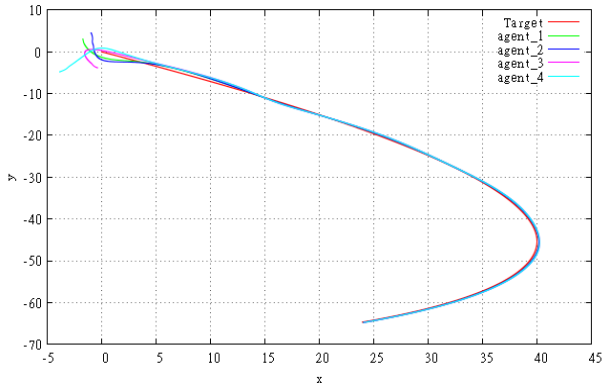


Figure 5. Trajectories of all the agents using LSM ($\gamma = 0$ in (18)) .

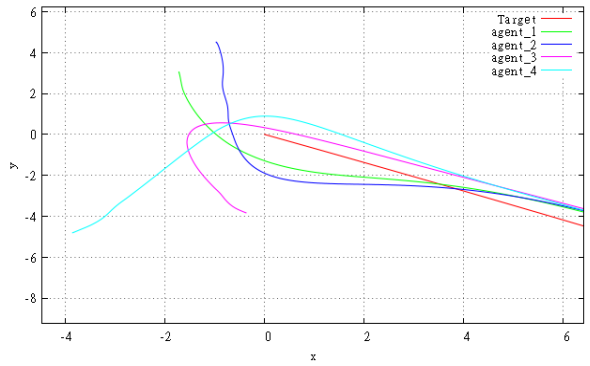


Figure 6. Transient trajectories of all the agents using LSM ($\gamma = 0$).

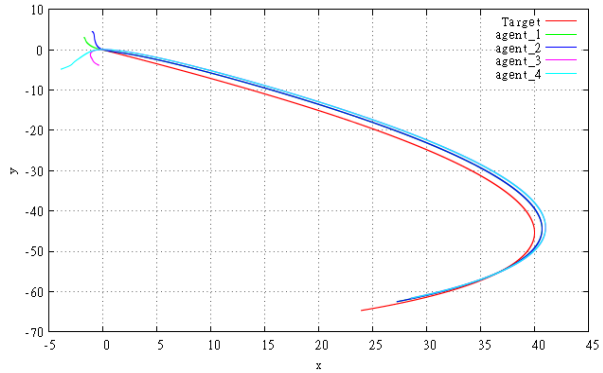
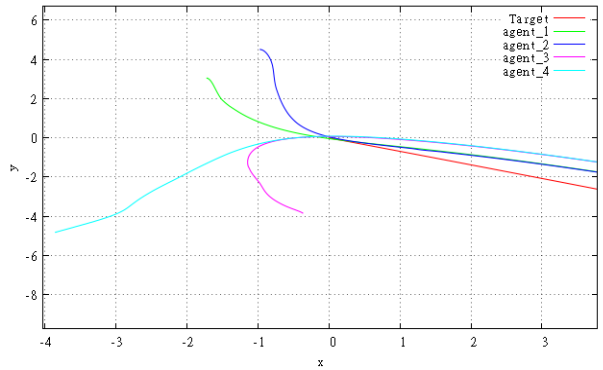


Figure 7. Trajectories of all the agents using TSM($\gamma = 1$ in (18)) .



\ Figure 8. Transient trajectories of all the agents using TSM ($\gamma = 1$).

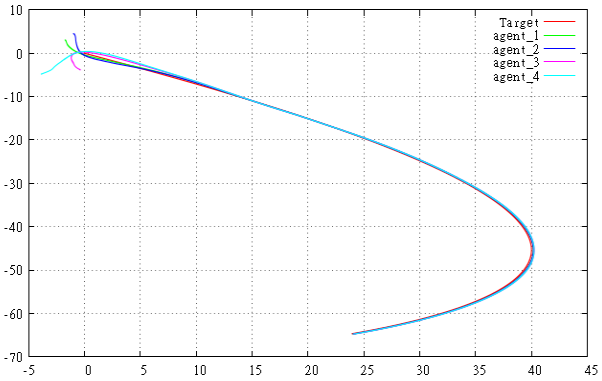


Figure 9. Trajectories of all the agents using TLSM.

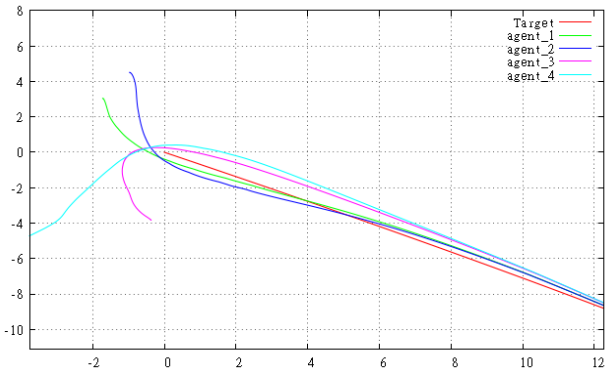


Figure 10. Transient trajectories of all the agents using T LSM.

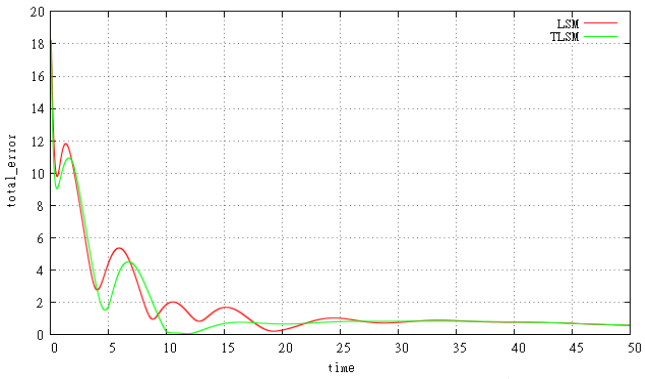


Figure 11. Trends in distances error between a leader and all follower agents.

Figures 9-10 show the results using our proposed LSM and TSM combined control, changing γ in (18) from 1.0 to 0.0 nonlinearly. Figure 11 shows trends in the consensus error between a leader and all follower agents for LSM and TLSM. From Figure 11, it can be found that consensus error by our proposed method is smaller than that of the conventional LSM method.

In order to confirm that our method is superior to the LSM method, we adopted the next performance index J, average error area;

$$J = \frac{1}{T}\sum_{t=0}^{T} e(t), \quad e = \sqrt{\sum_{i=1}^{4}\left\{\|x_L - x_{Fi}\|^2 + \left(\|y_L - y_{Fi}\|\right)^2\right\}}, \qquad (42)$$

where T means controlling time, $x_L = [x_l, \dot{x}_l]$, $y_L = [y_l, \dot{y}_l]$ mean position and velocity vectors in x direction and y direction of the leader, respectively. x_{Fi} and $\dot{x}_{Fi}$ mean those of the *i*th follower agent. e, J mean sum of consensus errors of between the leader and all the follower agents at time t, average consensus error area, respectively. Table 3 shows comparison of the control performances, that is, transient consensus error and average error area. Table 3 shows our proposed method is superior to the conventional LSM method

VIII. CONCLUSION

In this paper, we proposed the simple and useful method, that is, an adaptive LSM and TSM combined formation controller design method with auto- structuring fuzzy neural network. Additionaly, feature of the proposed method is that all the agents can exchange only information of positions according to the network structure of multi agents like Figure 4, making velocities of their virtual agents, and it has the variable structure to adapt for changes of the environment.

TABLE III. PERFORMANCE COMPARISONS OF BOTH METHODS.

Transient consensus error

agent	a_1	a_2	a_3	a_4	Total
LSM	0.148	0.161	0.234	0.24	0.78
TLSM (Proposed)	0.14	0.145	0.226	0.237	0.748

Average consensus error area

agent	a_1	a_2	a_3	a_4	Total
LSM	0.356	0.368	0.535	0.541	1.8
TLSM (Proposed)	0.316	0.321	0.478	0.489	1.604

REFERENCES

[1] Z. G. Hou, L. Cheng, and M. Tan, "Decentralized robust adaptive control for the multiagent system consensus problem using neural networks", IEEE Transactions on Systems, Man, Cybernetics-Part B: Cybernetics, vol. 39, no. 3, pp. 636-647.

[2] Q. Song, J. Cao, snd W. Yu, "Second-order leader-following consensus of nonlinear multi-agent systems via pinning control", Systems & Control Letters 59, 2009, pp. 553-562.

[3] L. Cheng, Z. G. Hou, M. Tan, Y. Lin, and W. Chang, "Neural-network-based adaptive leader-following control for multiagent systems with uncertainties", IEEE Transactions on Neural Networks, vol. 21, no. 8, 2010, pp. 1351-1358.

[4] M. Defoort, T. Floquet, A. Kokosy, and W. Perruquetti, "Sliding-mode formation control for cooperative autonomous mobile robots", IEEE Transactions on Industrial Electronics, vol. 55, no. 11, 2008, pp. 3944-3953.

[5] F. Morbidi and D. Prattichizzo,"Sliding mode formation tacking control of a tractor and trailer –car system", Robotics Science and Systems online proceedings, 2007.

[6] R. Cui, S. S. Ge, B. V. E. How, and Y. S. Choo, "Leader-follower formation control of underactuated autonomous underwater vehicles", Ocean Engineering, vol. 37, no. 17-18, 2010, pp. 1491-1502.

[7] W. Yu, G. Chen, and M. Cao, "Distributed leader-follower flocking control for multi-agent dynamical systems with time-varying velocities", System & Control Letters 5, 2010, pp. 543-552.

[8] Y. H. Chang, C. W. Chang, C. L. Chen, and C. W. Tao, "Fuzzy sliding-mode formation control for multirobot systems: design and implementation", IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics, vol. 42, no. 2, 2012, pp. 444-457.

[9] Chuan-Kai Lin,"Robust adaptive critic control of nonlinear systems using fuzzy basis function networks: An LMI approach", Information Sciences, 177, 2007, pp. 4934-4946.

[10] C. L. Chen, G. X. Wen, Y. J. Liu, and F. Y. Wang, "Adaptive consensus control for a class of nonlinear multiagent time-delay systems using neural networks", IEEE Transactions on Neural Networks and Learning Systems, vol. 25, no. 6, 2014, pp. 1217-1226.

[11] J. L. Lin, K. S. Hwan, and Y. L. Wan,"A simple scheme for formation Control based on weighted behavior learning ", IEEE Transactions on Neural Networks and Learning Systems, vol. 25, no. 6, 2014, pp. 1033-1044.

[12] X. Yu, M. Zhihong, Y. Feng, and Z. Guan, "Nonsingular Terminal sliding mode control of a class of nonlinear ", IFAC, 15th Triennial World Congress, Barcelona, Spain, 2002.

[13] A. M. Zou, K. D. Kumar, Z. G. Hou, and X. Liu, "Finite time attitude following control for spacecraft using terminal sliding mode and Chebyshev neural network", IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics, vol. 41, no. 4, 2011, pp. 950 - 963.

[14] Y. H. Chang, C. Y. Yang, W. S. Chan, C. W. Chang, and C. W. Taol, "Leader-following formation control of multi-robot systems with adaptive fuzzy terminal sliding-mode controller", IEEE International Conference on System Science and Engineering, 2013.

[15] M. Obayashi, Y. Yokoji, S. Uchiyama, L. B. Feng, T. Kuremoto, and K. Kobayashi, "Intelligent Following control method of a leader by Groups of agents with nonlinear dynamics", Proc. of 11th International Conference on Control, Automation and Systems, Korea, 2011.

[16] K. H. Cheng, "Auto-structuring fuzzy neural system for intelligent control", Journal of Franklin Institute, 346, 2009, pp. 267-288.

 ISBN: 978-1-61208-390-2

Are You Talking to Me?
Detecting Attention in First-Person Interactions

Luis C. González-García
and L. Abril Torres-Méndez

Robotics and Advanced Manufacturing Group
CINVESTAV Campus Saltillo
Ramos Arizpe, México
Email: carlos.gonzalez@cinvestav.edu.mx
abril.torres@cinvestav.edu.mx

Julieta Martinez, Junaed Sattar
and James J. Little

Department of Computer Science
The University of British Columbia
Vancouver, Canada
Email: {julm,junaed,little}@cs.ubc.ca

***Abstract*—This paper presents an approach for a mobile robot to detect the level of attention of a human in first-person interactions. Determining the degree of attention is an essential task in day-to-day interactions. In particular, we are interested in natural Human-Robot Interactions (HRI's) during which a robot needs to estimate the focus and the degree of the user's attention to determine the most appropriate moment to initiate, continue and terminate an interaction. Our approach is novel in that it uses a linear regression technique to classify raw depth-image data according to three levels of user attention on the robot (null, partial and total). This is achieved by measuring the linear independence of the input range data with respect to a dataset of user poses. We overcome the problem of time overhead that a large database can add to real-time Linear Regression Classification (LRC) methods by including only the feature vectors with the most relevant information. We demonstrate the approach by presenting experimental data from human-interaction studies with a PR2 robot. Results demonstrate our attention classifier to be accurate and robust in detecting the attention levels of human participants.**

Keywords–Human-robot interaction; Body pose classification; Least squares approximations; Raw range data analysis.

I. INTRODUCTION

Determining the attention of people is an essential component of day-to-day interactions. We are constantly monitoring other people's gaze, head and body poses while engaged in a conversation [1][2][3]. We also perform attention estimation in order to perform natural interactions [4][5]. In short, attention estimation is a fundamental component of effective social interaction; therefore, for robots to be efficient social agents it is necessary to provide them with reliable mechanisms to estimate human attention.

We believe that human attention estimation, particularly in the context of interactions, is highly subjective. However, attempts to model it have been relatively successful, *e.g.*, allowing a robot to ask for directions when it finds a human, as in the work of Weiss *et al.* [6]. Nonetheless, the state-of-the-art is still far from reaching a point where a robot can successfully interact with humans without relying on mechanisms not common to natural language. Recently, the use of range images to make more natural human-machine interfaces has been in the agenda of researchers, like in the case of the Microsoft Kinect™, which delivers a skeleton of a human that can be further used as a high-level feature of the human pose [7]. Although good results have been obtained with such devices in pose estimation, little effort has been devoted to further infer information about the user from such data. In this work, we use range data (similar to that shown in Figure 1) to infer the level of attention of the user, which is not explicitly given by the sensor output.

Figure 1. *Left*: Raw range input that a robot gets when trying to asses human attention, as described in this work. *Right*: Set-up scenario for our experiments. The PR2 robot approaches a human sitting at a desk..

Our approach is novel in that it uses raw depth images to evaluate the attention level of a subject, regardless of whether she is facing the depth sensor, in order to classify her pose in an attention scale. In this work, we focus on learning human attention from raw depth images by using the LRC algorithm, which can be exploited by social robots to determine the best moment to ask for support from a human sitting at her desk, like those found in common working or reading spaces.

The remainder of this paper is structured as follows: In Section II, we talk about how other authors have tackled the problem of attention awareness detection using images and range information as a source. In Section III, we describe the problem that this paper faces, attention estimation from a first person perspective using only raw range information. In Section IV, we walk through the technical aspects of the methodology that we propose (LRC). In section V, it is described the actual set-up and execution of the experiments, as well as the interpretation and discussion of the data gathered from them. Finally, in Section VI, our conclusions and suggestions for future work are exposed.

II. RELATED WORK

The problem of attention awareness detection, despite its relevance, remains largely unexplored in the HRI literature. Here we present some of the building blocks of our work.

 ISBN: 978-1-61208-390-2

A. Pose, Head and Gaze Estimation

Some of the most effective social cues for attention estimation are gaze, body and head poses. Fortunately, a large body of knowledge has been gathered in these areas.

Shotton *et al.* [7], used a single Red-Green-Blue+Depth (RGB+D) camera to perform pose estimation and body parts recognition. A randomized decision forest was trained on synthetic data that covered a wide range of human poses and shapes. Features were obtained by computing the difference of depth between two points. A further speedup was achieved by providing a GPU implementation. Vision-only approaches range from the Flexible Mixtures-of-Parts [8], an extension of the Deformable Parts Model [9] which explicitly accounts for different body deformations and appearances, to leverage poselet-based part detections for further constraining optical-flow-based-tracking of body parts [10].

Head pose estimation can be seen as a sub-field of full-body pose estimation. In fact, Kondori, Yousefi, Haibo and Sonning [11] extended the work of Shotton *et al.* to Head Pose Estimation in a relatively straightforward manner. Similarly, the problem becomes much harder when depth information is no longer available.

For a more in-depth treatment of the subject, as well as for recent advances in gaze estimation, we direct the reader to the reviews made by Murphy-Chutorian and Trivedi [12] and Hansen and Qiang [13].

B. Awareness Detection in Computer Vision

Estimating attention from visual input has been studied particularly in the context of driving. Doshi and Trivedi [14] built a system that incorporated cameras observing both the human subject and her field of view. By estimating the gaze of the subject and the saliency map from her viewpoint, they used Bayes' rule to obtain a posterior distribution of the location of the subject's attention. Our work is different from theirs since just as in person-to-person interactions, we do not have access to the field of view of the person, but we might rather *be* a part of it.

Also related are Mutual Awareness Events (MAWEs). MAWEs are events that concentrate the attention of a large number of people at the same time. In this context, Benfold and Reid [15] built upon evidence from the estimated head poses of large crowds to guide a visual surveillance system towards interesting points.

C. First-Person Interaction

Recently, some work has been devoted to transfer knowledge gained from third to first person perspectives. Ryoo and Matthies [16] performed activity recognition from a first-person viewpoint from continuous video inputs. They combined dense optical flow as a global descriptor and cuboids [17] as local interest point detectors, then built a visual dictionary to train an SVM classifier using multi-channel kernels.

D. Human Attention and Awareness Estimation

To this day, human attention remains an active area of research. A widely accepted model of attention was proposed by Itti and Koch [18], where attention is understood as the mixture of "bottom-up", *i.e.*, unconscious, low-level features of an image, and "top-down", *i.e.*, task-oriented mechanisms that the subject controls consciously. Later work by Itti and Baldi incorporated the element of Bayesian surprise [19], *i.e.*, which states things that are different on the temporal domain attract attention, but with time they get incorporated into our world model and become less relevant. We keep this factor in mind when designing the experiment, as people who are not used to interacting with a robot might direct their attention to it just because it is something new, rather than because of its actions.

It is also important to mention that in order to attract the user's attention, the robot has to be attentive to the person. This often involves mimicking human mechanisms that indicate attention. Bruce, Nourbakhsh and Simmons [4] found that if a robot turns its head to the person whose attention it wants, then the probability of the person cooperating with the robot is greatly increased. This is also exploited by Embgen *et al.* [20], who further concluded that the robot needs only move its head to transmit its emotional state. Nevertheless, no further analysis was performed to determine whether or how this robot-to-human non-verbal communication impacts HRI.

E. Linear Regression Classification

LRC is a simple yet powerful method for classification based on linear regression techniques. Naseem, Togneri and Bennamoun [21] introduced LRC to solve the problem of face identification by representing an image probe as a linear combination of class-specific image galleries. This is performed by determining the nearest subspace classification and solving the inverse problem to build a reconstructed image, choosing the class with the minimum reconstruction error. During the training phase, the inputs are added to the database using a greedy approach. Every input image is required to add a minimum information gain in terms of linear subspace independence: only if they fulfill the criterion, they are added to the database. This keeps the database size small, and allows for efficient training and classification. To the best of our knowledge, we are the first to apply this method to raw depth image data.

III. Problem Formulation

The problem that we address is human attention estimation from a first-person perspective. At a coarse level, we define attention in three categories: a) *null* attention, b) *partial* attention and c) *total* attention. We believe that this simple scale is enough to model a wide range of situations, since they encode the willingness of a user to engage in interaction. If robots are meant to be efficient social agents, it is imperative to be able to detect the right instance to start, maintain and end task-oriented interactions with humans.

A. Scenario

In our study, we assume that the robot wants to interact with a human who is sitting at a desk, a common occurrence in an office environment (see Figure 1); the robot wants to start an interaction approaching the human from one side. The situation is analog to a human approaching a coworker at her desk, willing to know if she is available for a given task. For our experiments, we use the Willlow Garage PR2 robot, with users occupying lab workstations with computer terminals. Having the experiment occur within the confines of the lab spaces ensures users' attentions are not unduly attracted to the robot, as having the robot around is a fairly common occurrence in the lab.

 ISBN: 978-1-61208-390-2

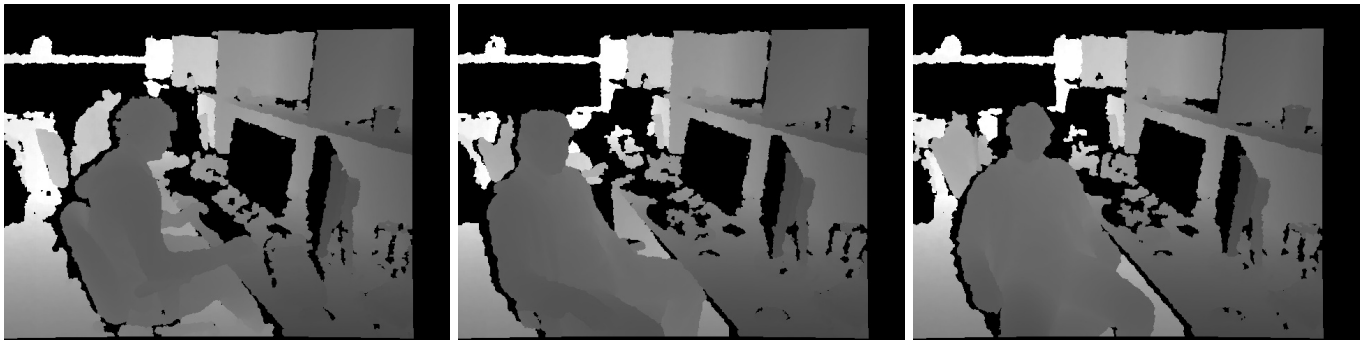


Figure 2. Representative raw-depth images of the three levels of attention. From left to right, *null* attention, *partial* attention and *full* attention. The images were captured using a Kinect™ mounted on the PR2 head.

B. Data

In order to evaluate human attention and obtain the best moment to ask for support, the robot relies only on a set of depth images captured with its Kinect™. For initial tests, the data was captured using a separate Kinect™ sensor mounted on a tripod simulating the pose that the sensor would have above the PR2. For our study, the data captured consisted only of depth information, allowing our approach to be robust against illumination variations, as well as other appearance changes. The intention is to demonstrate that our approach is robust enough so that visual information is not required, and efficient enough to run on a constrained computational platform while performing in real-time.

To build the training database, a subject is seated at a desk and asked to perform activities that simulate the three levels of attention of our scale. We describe each attention levels next:

1) *Null attention (class 1):* the subject's posture is such that she is facing the computer monitor, pretending she is busy, working;
2) *Partial attention (class 2):* the subject's posture is such that she is not facing the monitor, nor the robot, but rather facing somewhere in between;
3) *Full attention (class 3):* the subject's posture is such that she is facing the robot.

The subject is free to simulate the three attention levels according to her discretion, as long as the basic guidelines described above are satisfied. We recorded the movements of the subject and repeated the experiment several times; each one by placing the depth sensor in different configurations (*i.e.*, changing position, elevation and orientation), allowing for a more versatile training set. Examples of the range data are shown in Figure 2.

IV. Technical Approach

Our approach towards determining attention levels consists of an offline training stage and an online detection stage. The training step includes capturing depth snapshots (or video streams) of the user at her workstation or desk, extracting features and constructing class-specific feature matrices to build an attention classifier. During attention classification, instantaneous depth image snapshots from the Kinect™ are fed into the classifier, and the class with the minimum linear independence with respect to the training data is chosen as the likely attention level. The following sections provide technical details of these individual steps.

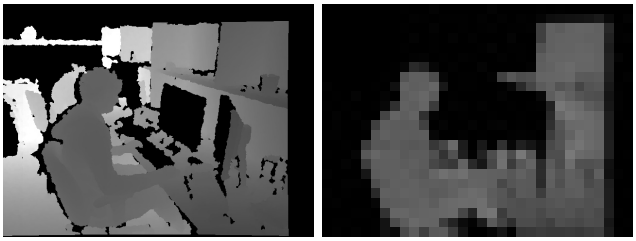


Figure 3. *Left*: Raw depth image and *Right*: preprocessed image ($\gamma = 1/20$ and distance = 2 meters).

A. Interaction Setting

The HRI is carried out as follows. First, the robot approaches and stands on either side of the human, using the range data to assess if the human is occupied, and if that is the case, continue evaluating the best moment to ask for support. If the human remains busy for an extended duration, then the robot does not engage in interaction and attends to other tasks. The aim is to evaluate if a robot standing close to a human working at a desk can accurately estimate the degree of attention of the human, and use this information to ask for support at the correct time instance. Data for our training set is obtained from the Kinect™ mounted on the robot in such scenarios, and is limited to range data only. Collected range images consist of particpants performing actions corresponding to the attention levels that we defined. The image sensor is placed on both sides of the user (see Figure 4), while recording the actions of the subject.

B. Features

For our LRC, the features consist of depth image data downsampled to $\gamma = \frac{1}{20}$ scale (see Figure 3), and reshaped by concatenation of its columns, similar to the methodology of Naseem, Togneri and Bennamoun [21]. However, as we are working with depth images, and we do not want the scene background to interfere with the learning and classification processes, the depth images are preprocessed to remove unwanted data. Specifically, we consider only those depth values up to a specific range, which accounts for the approximate physical distance between a robot and a human sitting at his desk under the current interaction scenario. This distance was empirically observed to be approximately 2 meters from the Kinect™ sensor.

 ISBN: 978-1-61208-390-2

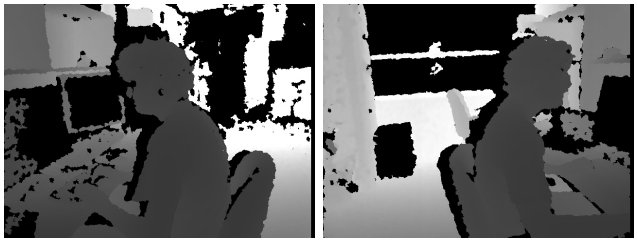


Figure 4. Range data from the left and right profile of the subjects. Both were included on the database for this study.

C. Training

Naseem, Togneri and Bennamoun [21] consider a database with photographs of people's faces. This is translated into a small number of sample images per class (subject), avoiding the time constraint of solving the pseudoinverse involved on a linear regression, experienced on large datasets, like videos. The novelty of our work is that we deal with this constraint (big datasets) by analyzing the linear independence of the images. By doing this, we dismiss all the new pictures that does not add relevant information to the LRC. Thus, we can condense the dataset into representative images, without losing relevant information. This allows us in turn to achieve an efficient LRC in real time.

Let $\mathrm{Y} \in \mathbb{R}^m$ be a vector of a given matrix $\mathrm{X} \in \mathbb{R}^{m \times n}$. We used a linear regression technique to analyze the linear independence of Y, as shown in Algorithm 3. In this algorithm, vector Y is projected onto the column space of X, then an error is calculated by subtracting the resultant projected vector Y_c to the original vector Y, giving as result the projection error ε. This ε is a metric that is used to measure the linear independence of Y with respect to the column space of X.

D. Algorithm

The overall algorithm is divided into two parts,

1) *Build X:* This procedure (Algorithm 1) analyzes the range image database and builds one matrix *per* attention class that contains the most significant collection of images in that class, and ensures a maximum degree of linear independence between images in the same class.
2) *Classify Y:* This procedure (Algorithm 4) is responsible for using the class matrices generated by the *Build X* algorithm, as well as the input depth image, to classify that image into one of the known classes. It also outputs the projection error of the image with respect to the column space generated by each of the class matrices, choosing the class with the minimum projection error.

It is important to mention that in order to reduce the overhead of a pseudo inverse calculation, $\mathrm{X}_i^\dagger$ is calculated only when X_i changes, thus X_i and $\mathrm{X}_i^\dagger$ are saved and computed only once for classification.

Algorithm 1 Build X, the probe database.

Require: Threshold τ
for each class i **do**
  Img $\leftarrow$ Random unseen image. ▷ Initialize X_i
  Y $\leftarrow$ FEATURES(Img)
  Append Y to X_i
  Compute $\mathrm{X}_i^\dagger$ ▷ Build X_i
  for each new image nImage **do**
    Y $\leftarrow$ FEATURES(nImage)
    $\varepsilon \leftarrow$ LINEARINDEPENDENCE$(\mathrm{Y}, \mathrm{X}_i, \mathrm{X}_i^\dagger)$
    if $\varepsilon \geq \tau$ **then**
      Append Y to X_i
      Compute $\mathrm{X}_i^\dagger$
    end if
  end for
end for
save($\mathrm{X}_i^\dagger$, X)

Algorithm 2 Feature extraction. Performs downsampling and reshaping.

procedure FEATURES(Image, γ)
  Cut the image background.
  Down-sample the image by a factor γ.
  Reshape the image to a column vector.
  return The post-processed Image.
end procedure

V. Experimental Results

We conducted a number of trials to evaluate our proposed approach. To train our system, we used range data of the three specific attention classes from 5 different participants, following the process described in Section IV. For each participant in the training process, we obtained video streams for three different attention levels, in two different Kinect™ positions; each of the video streams have dimensions of 640×480 pixels, and have approximately 500 frames each. This resulted in a total of $15,000$ frames for the training process. The attention matrices X_i have average dimensions of 39×768, with 39 images downscaled to $\frac{1}{20}$ of their original dimensions. The value of the threshold τ was empirically set at 4.0 for all trials. Training and classification was performed on a PC with an Intel Core-i5™ processor running at 1.7 GHz, with 2GB of memory and under the Ubuntu 12.04 Long-term Release (LTS) edition. The code was implemented in C++ using the Robot Operating System (ROS) C++ bindings.

Our results are summarized in Figures 5 and 6. The figures show frame-by-frame reconstruction errors of a test video. The errors represents the linear independence of an input image with respect to the column space of each class-specific database X_i, $i \in \{1, 2, 3\}$.

Figure 5(a) shows the classification of a video that was used during the training phase, as expected, the projection error is close to zero almost all the time. When the error reaches zero, is an indication that the current image passed the linear independence test, and it was used on the learning phase. While this is not illustrative of the accuracy of our algorithm, it does illustrate the fact that most of the information used during training is redundant, and that by keeping only a small fraction of it we can achieve a low reconstruction error. Figure 5(b) shows the classification performance on a video sequence that was not used during the training phase. While the reconstruction error is larger in this case, it is nonetheless sufficient to perform classification accurately. The key observation is that irrespective of the actual reprojection error numbers, there is clear separation between the actual attention class errors and the errors of the other attention classes, which leads to distinct identification of the user's attention level.

 ISBN: 978-1-61208-390-2

Algorithm 3 Measure the linear independence of Y with respect to database X.

Require: $Y \in \mathbb{R}^m$, $X, \in \mathbb{R}^{m \times n}$, $X^\dagger \in \mathbb{R}^{n \times m}$
1: **procedure** LINEARINDEPENDENCE(Y, X, $X^\dagger$)
2: $Y_c \leftarrow XX^\dagger Y$ ▷ Reprojection of Y.
3: $\varepsilon \leftarrow ||Y_c - Y||^2$
4: **return** ε ▷ The reprojection error is the metric.
5: **end procedure**

Algorithm 4 Classify a new vector Y.

Require: An input Image. Precomputed X_i and $X_i^\dagger$ for each class i, the class-specific databases and their pseudo-inverses.
1: **for** each class i **do**
2: $Y \leftarrow$ FEATURES(Image)
3: $\varepsilon_i \leftarrow$ LINEARINDEPENDENCE$(Y, X_i, X_i^\dagger)$
4: **end for**
5: **return** $\text{argmin}_i(\varepsilon_i)$ ▷ Return the class with minimum error.

The LRC is done *per* frame, by choosing the minimum of these errors, and using a leave-one-out validation process during training (*i.e.*, while building X, one subject was left out the matrix). Hence, as is observed in Figure 5, the robot is capable of correctly estimating the attention level, even when testing on a subject that was not included on the database. Figure 7 shows the difference in poses between the training (left) and testing (right) sets.

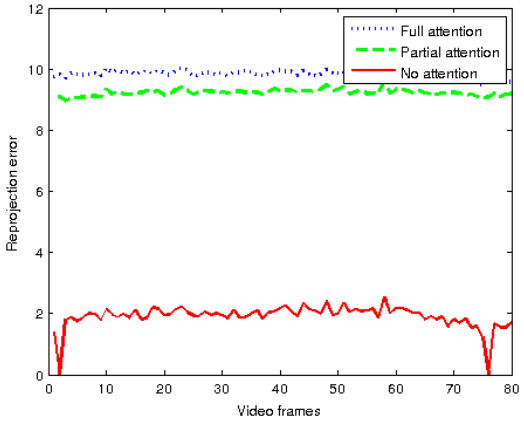
Nevertheless, large variations in the RGB+D sensor pose can lead to reduced performance of our algorithm. This is demonstrated in Figure 6, where the robot (and thus the Kinect™) is continually placed in positions not used to capture training data, while the system tries to detect a user in Class 3 (*i.e.*, full) attention level. As the Kinect™ changes its pose, the errors levels vary, resulting in an inaccurate classification. Note that the seperations between classes on the error scale are also reduced, resulting in degraded accuracy. The slopes on the figure represent displacement of the Kinect™.

A. Quantitative Analysis

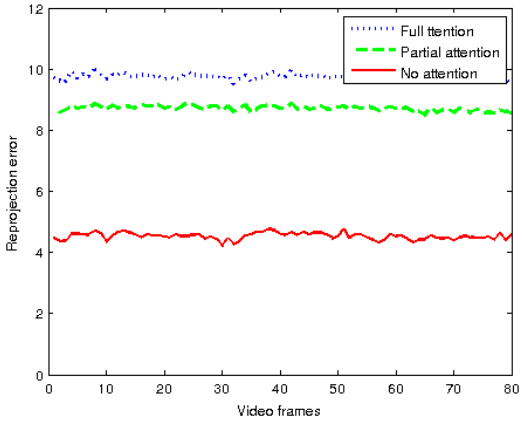
In order to evaluate and validate our proposed method, we compare it against other common approaches for estimating visual attention based only on visual information, namely the Head Pose and Gaze attention estimation [22][14]. Ideally the comparison should be carried out using a ground truth of the subject attention, but this is extremely subjective, due to the inherently complexity of the human behavior, for this reason a simulated labeled attention is used as ground truth. To simulate this baseline, attention is lurked to the camera by showing interesting images to the user in a display beneath the sensor; similarly the attention is also directed outside the camera showing interesting images in another display.

For the purpose of this evaluation, the three attentive states are wrapped into two main states, attention or no attention towards the sensor, so, when the estimator and the simulated ground truth coincide on the attentive state of the user, a 1 or *OK* is recorded, otherwise a 0 or *NOT OK* is recorded. In the end, all this records are averaged in order to obtain an average accuracy of the estimator against the simulated ground truth, which in the context of this paper, it can be used as the estimator's main metric of performance.

The performance comparison is summarized on the Table I.



(a) Included in the dataset



(b) Unkown subject

Figure 5. Reprojection class error vs. Time. An input video from Class 1 (no attention) is being classified in real time. *(a)* LRC over a dataset that was included on the learning database. *(b)* LRC over data that was not included on the learning database.

TABLE I. Comparison of proposed Raw-Range-Information attention estimator with approaches based purely on visual information, Head Pose and Eye Gaze (coarse). Avg. Accuracy corresponds to the attention state between that reported by the estimator and the labeled attention.

Estimator	*Avg. Accuracy*
HeadPose-based	0.7333
EyeGaze-based	0.5667
LRC on raw range (proposed)	**0.8500**

VI. Conclusions and Future Work

We have presented a novel method to accurately estimate the attention of a person when interacting with a robot.

The results demonstrate that our method outperforms other models for attention estimation based solely on visual information (Table I).

The estimation is performed using raw depth data of the person's body pose. We use a LRC with the selection of the best linearly independent depth-images during the training phase. Our method also effectively discards redundant information during the training phase, while maintaining good performance on previously-unseen sequences. In addition to being completely independent from appearance and illumination changes, our approach is robust to small pose variations,

 ISBN: 978-1-61208-390-2

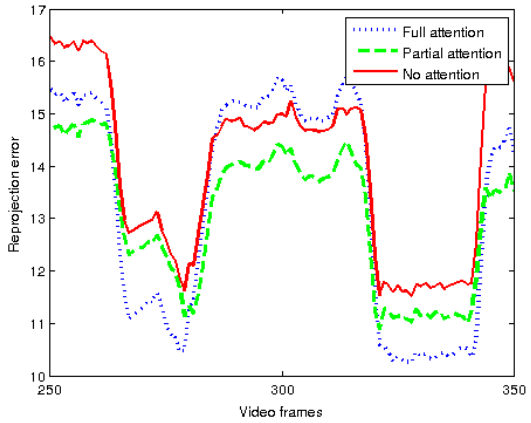


Figure 6. Effect of sensor displacement on classifier performance (Full attention).

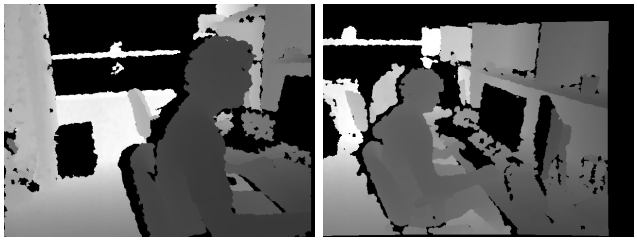


Figure 7. Difference in the position of the range sensor between images that were included (*left*) and not included (*right*) on the database for this study.

from the training data. The main advantages of our classifier are its simplicity, real-time computability and small memory footprint, which is ideal for implementation on board robots for man-machine interaction tasks.

In future work, we intend to explore attention estimation in a wider variety of settings, perform larger-scale experiments (encompassing more attention classes, more subjects, more situations), and explore the limits of the LRC approach. Particularly, we plan to fuse this approach with other face and gaze detectors in order to achieve a short-long distance attention estimator.

ACKNOWLEDGEMENTS

We would like to thank CONACyT and CBIE, under the ELAP, as well as Fonds de Recherche du Quebéc – Nature et Technologie, for their support and project funding.

REFERENCES

[1] M. Argyle, Bodily communication. Routledge, 2013.

[2] M. Knapp, J. Hall, and T. Horgan, Nonverbal communication in human interaction. Cengage Learning, 2013.

[3] N. Hadjikhani, K. Kveraga, P. Naik, and S. P. Ahlfors, "Early (n170) activation of face-specific cortex by face-like objects," Neuroreport, vol. 20, no. 4, 2009, p. 403.

[4] A. Bruce, I. Nourbakhsh, and R. Simmons, "The role of expressiveness and attention in human-robot interaction," in Robotics and Automation, 2002. Proceedings. ICRA '02. IEEE International Conference on, vol. 4, 2002, pp. 4138–4142.

[5] S. Lang, M. Kleinehagenbrock, S. Hohenner, J. Fritsch, G. A. Fink, and G. Sagerer, "Providing the basis for human-robot-interaction: A multi-modal attention system for a mobile robot," in Proceedings of the 5th International Conference on Multimodal Interfaces, ser. ICMI '03. New York, NY, USA: ACM, 2003, pp. 28–35. [Online]. Available: http://doi.acm.org/10.1145/958432.958441

[6] A. Weiss, J. Igelsbock, M. Tscheligi, A. Bauer, K. Kuhnlenz, D. Wollherr, and M. Buss, "Robots asking for directions - the willingness of passers-by to support robots," in Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on, March 2010, pp. 23–30.

[7] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-time human pose recognition in parts from single depth images," in Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, June 2011, pp. 1297–1304.

[8] Y. Yang and D. Ramanan, "Articulated pose estimation with flexible mixtures-of-parts," in Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, June 2011, pp. 1385–1392.

[9] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, June 2008, pp. 1–8.

[10] K. Fragkiadaki, H. Hu, and J. Shi, "Pose from flow and flow from pose," in Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, June 2013, pp. 2059–2066.

[11] F. Kondori, S. Yousefi, H. Li, and S. Sonning, "3d head pose estimation using the kinect," in Wireless Communications and Signal Processing (WCSP), 2011 International Conference on, Nov 2011, pp. 1–4.

[12] E. Murphy-Chutorian and M. Trivedi, "Head pose estimation in computer vision: A survey," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 31, no. 4, April 2009, pp. 607–626.

[13] D. Hansen and Q. Ji, "In the eye of the beholder: A survey of models for eyes and gaze," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 32, no. 3, March 2010, pp. 478–500.

[14] A. Doshi and M. Trivedi, "Attention estimation by simultaneous observation of viewer and view," in Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on, June 2010, pp. 21–27.

[15] B. Benfold and I. Reid, "Guiding visual surveillance by tracking human attention," in Proceedings of the 20th British Machine Vision Conference, September 2009, pp. 1–11.

[16] M. Ryoo and L. Matthies, "First-person activity recognition: What are they doing to me?" in Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, June 2013, pp. 2730–2737.

[17] P. Dollar, V. Rabaud, G. Cottrell, and S. Belongie, "Behavior recognition via sparse spatio-temporal features," in Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 2005. 2nd Joint IEEE International Workshop on, Oct 2005, pp. 65–72.

[18] L. Itti and C. Koch, "Computational modelling of visual attention," Nature reviews neuroscience, vol. 2, no. 3, 2001, pp. 194–203.

[19] L. Itti and P. Baldi, "Bayesian surprise attracts human attention," Vision research, vol. 49, no. 10, 2009, pp. 1295–1306.

[20] S. Embgen, M. Luber, C. Becker-Asano, M. Ragni, V. Evers, and K. Arras, "Robot-specific social cues in emotional body language," in RO-MAN, 2012 IEEE, Sept 2012, pp. 1019–1025.

[21] I. Naseem, R. Togneri, and M. Bennamoun, "Linear regression for face recognition," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 32, no. 11, Nov 2010, pp. 2106–2112.

[22] A. Doshi and M. Trivedi, "Head and gaze dynamics in visual attention and context learning," in Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on, June 2009, pp. 77–84.

 ISBN: 978-1-61208-390-2

A Proposed Method to Support Awareness of Specialization for Interdisciplinary Communication Education

Tadashi Fujii, Kyoko Ito, Shogo Nishida
Department of Engineering Science
Graduate School of Engineering Science Osaka University
Osaka, Japan
Emails: {fujii@nishilab.sys.es.osaka-u.ac.jp, ito@sys.es.osaka-u.ac.jp, nishida@sys.es.osaka-u.ac.jp}

***Abstract*—The more technologies are being developed and used, the more problems occur in many technical fields. Interdisciplinary communication is suitable to solve such problems. However, the education fostering this approach is still insufficient. The purpose of this study is to develop an interface to support the awareness of specialization so that we provide an effective interdisciplinary communication education for participants. Towards this goal, we utilize the transition of speakers' specializations following turn-taking and we define it as "patterns". To support awareness of specialization, this study focuses on the rate at which participants recognize other experts' specialization and specify the patterns when the awareness of specialization is easy to obtain. Experiments have been performed on 16 participants and were analyzed based on quantity and accuracy of the information to ascertain the effect of the proposed pattern. Our results show that the proposed pattern can give information at the rate of about 50% and support participants' recognition of other experts' specialization. From this result, the proposed pattern can support participants' recognition of other experts' specialization. Based on this, we propose an interface that shows when proposed patterns are likely to create chances to perceive specialization.**

Keywords–Interdisciplinary; Communication; Specialization.

I. INTRODUCTION

In recent years, experts' fields of science technology have become increasingly subdivided into many areas. By separating fields, we can encourage the growth of specific disciplines. However, some problems accompany the subdivision of learning fields; for example, we miss out on a holistic picture and it becomes more difficult to solve problems which span more than one technical field. To solve these problems, interdisciplinary communications are becoming more important [1]. For example, in the field of human / computer interaction, people are developing systems to improve interaction, not only between humans and computers, but also among humans. When we develop such systems, it requires the knowledge of psychology and sociology to perform practical interactions. In addition, there are various subjects such as medical science and economics to be considered. If we focus on such fields, we should not only investigate relevant studies, but also communicate with experts in those fields. However, there are few opportunities to communicate with other fields in a practical way in the Japanese educational environment. For that reason, it is thought that interdisciplinary communication education is insufficient [2].

In such a situation, interdisciplinary discussions are focused as a method of learning interdisciplinary communication [3], because interdisciplinary discussion may function as training for communication with people who have different specializations. In learning methods by using discussion, it is important to have interests in the differences between specializations and to recognize these features. By examining other expert's specialization, we better understand our own research position in academia; by acquiring different experts' knowledge and viewpoints, we have the opportunity to develop new technologies. However, it is difficult to fully comprehend other participants' specialization in actual interdisciplinary communication. These difficulties are due to the complexity of the contents of interdisciplinary communication that have diverse participants' specialization. Moreover, participants must voice their ideas in discussions; therefore, they cannot focus fully on paying attention to others' specialization. In the field of cognitive psychology, it is known that people cannot perceive anything which they do not understand [4]. This way, it is insufficient for people who are not familiar with interdisciplinary communication to learn it merely by discussion with experts who have different specialization.

The purpose of this research is supporting the awareness of specialization for effective learning of interdisciplinary communication. For that reason, we propose a method to support awareness of specialization. We utilize the transition of speakers' specialization following turn-taking as "patterns", and examine a method to determine the patterns that make it easier to comprehend other participants' specialization. We ascertain whether the patterns can help participants to recognize the specialization through experiments. Based on the result of an experiment, we propose an interface that shows the timing when specialization is apt to be recognized to help participants recognize the specialization. The rest of the paper is organized as follows. In Section II, a review of the literature on related work and the viewpoint of this study are given. In Section III, we define the elements utilized in this study and propose the information to support the awareness of specialization. In Section IV, we describe the experimental method and the result. In Section V, the proposal of interface idea is given. In Section VI, a conclusion and future works are given.

 ISBN: 978-1-61208-390-2

II. RELATED WORKS

A. Related works

In recent years, many studies have investigated interdisciplinary communications. Fujigaki ascertained that multi specialized knowledge caused difficulties in understanding each other in interdisciplinary communication [5]. She defined the axis classifying scientific ideas, analyzed them and proposed a practical method of knowledge integration to remove the difficulties. Visualization methods are proposed by many studies to support interdisciplinary communication. Sumi et al. developed an interface denoting background knowledge that each participant's specialization had on a computer display [6]. They showed it could solve the problems caused by distance of knowledge. Huub et al. offered ontological knowledge base and Modeling Support Tools to support the multidisciplinary modeling process for water management that sometimes lacks mutual understanding between modeling team members [7]. They showed that tools can facilitate cooperation in teams. Lisa et al. visualized a cluster that showed studies' relationships in interdisciplinary fields [8]. This study suggested a system intended to reveal each study's position and its outline.

We refer to studies about turn-taking. Turn-taking is a speakers' exchange system defined by Sacks et al [9]. There are many studies that use turn-taking to support discussion. Cao et al. showed that people could smoothly shift the right to speak by obtaining feedback on turn-taking [10]. Dimicco et al. created an application that visualizes information about turn-taking and speaking time [11]. From the investigation, visualization can help participants reconsider and better understand their interactions from social and behavioral viewpoints.

B. View point

In this research, we support recognition of specialization in the discussion for learning interdisciplinary communication. Therefore, we consider which elements of discussion we should use to support awareness. In discussion, specialization generally appears in participants' utterances. Moreover, how to represent each specialization and the contents thereof depends on the speakers' field of academic specialization. We define the learning domain that discussion participants specialize in as their "field." For these reasons, we focus on participants' utterances and fields.

In addition, we must also consider the discussion's features because we selected a learning method that uses dialog. Discussion consists of interactive utterances by participants. In this respect, it is the same as interdisciplinary communication. For that reason, we should study the features relevant to awareness of specialization from interaction with more than one participant. Therefore, we focus on turn-taking with regards to recognizing the characteristics of specialization.

We mentioned turn-taking and fields as elements to observe. We connect them and treat the transition of speakers' specialization following turn-taking. We defined the transition as a "pattern" and we use it to specify the timing that governs when participants are easily able to recognize the specialization. We describe an example of the pattern. By denoting humanities experts as "H" and science experts as "S", if H speaks just after S, we form the pattern "SH."

C. Difference from existing research

In existing research, many methods are proposed for supporting awareness of specialization. However, these studies do not refer to participants whose understanding of interdisciplinary communication is insufficient. There is some possibility that if those people used the proposed systems, they could not communicate smoothly and make a discovery through awareness of specialization. It is not until those people are educated that they can use the existing systems. With regard to studies that focused on turn-taking, those proposed methods are not for learning but rather for interaction support. This study treats turn-taking as one element to support the awareness of specialization in group discussions necessary for interdisciplinary communication within academia. That distinguishes our study from existing studies in focusing on basic learning and using it for interdisciplinary communication.

III. TOWARD THE SUPPORT OF THE SPECIALIZATION AWARENESS

A. Awareness of specialization and turn-taking

It is necessary to define what constitutes "awareness of specialization" for the purpose of interdisciplinary communication support. Specialization is composed of education and knowledge that participants have attained; therefore, the features of specialization are influenced by which disciplines the participants have learned. When participants attempt to recognize other experts' specializations, they compare their knowledge with other experts' utterances. Therefore, what participants can recognize is whether the utterance has specialization from the viewpoint of their own specialization. For that reason, this study treats "awareness of specialization" as "recognizing the features of other specializations,"

We are required to consider a relationship between fields and specialization because we are focusing on transition between fields. In Japan, academic fields are often classified into two types: "Humanities" and "Sciences" [12]. Generally, this classification is based on universities' departments or high schools' courses; however, it is difficult to delineate a clear division. Moreover, there are various specializations, and their classification and each participant's understanding of the classification is different. For that reason, each specialization should be labeled based on its field to avoid confusion.

 ISBN: 978-1-61208-390-2

Sacks et al. defined turn-taking as consisting of five rules [9][12]. However, his rules did not stipulate that participants begin to speak simultaneously or while other people are speaking. Schegloff [13] and Jisun [14] defined how turns should be treated when such cases arise. Accordingly, this study combines Sacks' Schegloff's and Jisun's rules and defines turn-taking as having seven rules.

This study focuses on awareness of specialization and pattern; therefore, it is necessary to consider these relationships. It is expected that there are easily recognized patterns because specialization necessarily reflects participants' disciplines and their awareness is based on the comparison of their own paradigm to other experts' specializations.

B. Proposed method

This study defines a "specialization rate" as the rate at which participants recognize other experts' specialization. We estimate the timing which participants will recognize the specialization and inform them about the timing to encourage awareness. For example, if a certain pattern appears x times and specializations emerge at y times just behind that pattern, then specialization rate " α " is defined as in the numerical expression below.

$$\alpha = y/x \quad (1)$$

It can be expected that we can recognize patterns more easily if the patterns are decided by specialization rate. We propose a method to obtain such patterns. Specifically, we classify patterns based on each speaker's field, derive each specialization rate, and form a pattern from fields that have the highest specialization rates. This study defines the pattern obtained by the above method as the "specialization pattern." Moreover, we define the patterns that are same as specialization pattern except one field as the "proximate patterns." This study calls for greater attention to awareness of specialization in subsequent utterances by presenting these patterns.

We illustrate this point with an example deriving a pattern composed of three utterances and two fields of Humanities and Sciences. Hereinafter, Humanities are referred to as "H" and Sciences are referred to as "S." If there are N utterances in the discussion, there are N-2 patterns in the same discussion. At first, we divide these patterns into two types based on first speaker's field whether H or S. Second, we calculate each case's specialization rate and select the field which has the higher specialization rate. In the same way, we compose the specialization pattern for the second and third field. Finally, we define proximate patterns from the decided pattern. For example, if we get pattern SSH, the proximate patterns are defined as HSH, SHH, and SSS.

IV. EXPERIMENT

A. Purpose

The purpose of the experiment is to gain information about turn-taking and awareness of specialization and examine the effect of specialization pattern. Specifically, we derive specialization pattern and proximate patterns and confirm the rate of the utterance just behind the patterns containing specialization. From the result, we propose an interface.

Figure 1. Experiment

B. Abbreviations and Acronyms

The purpose of this research is to support recognizing specialization for interdisciplinary communication learning; therefore, it is necessary to observe interdisciplinary discussions. Participants of the experiment should have a certain amount of knowledge and little experience of interdisciplinary communication because the candidates for support are unfamiliar with interdisciplinary discussion. The classification of specialization uses "Humanities" and "Sciences" which are referred to as "H" and "S" to prevent confusion. H and S are classified based on Universities' departments which are familiar in Japan because it is said that the factor which determines to which fields a person belongs to is the relevant university's departments [15][16]. We adopt current topics as the subject for the discussion because it seems more likely that participants know about the topics and they have original knowledge of their fields. It is necessary to obtain information about turn-taking and awareness of specialization while participants engage in discussion. However, participants are unfamiliar with interdisciplinary discussion; it is difficult for them to carry on a discussion and recognize the specialization at the same time. This study obtains information about turn-taking and awareness of specialization separately. Participants are able to focus on discussion while they speak and concentrate on awareness of specialization while they try to recognize it in order to get valid data.

C. Experimental method

This study carries out an experiment in accordance with the following conditions.

- Subject: Should we retain nuclear power plants in Japan?
- Participants: 16 University students (14 males, 2 females: early twenty years old)
- H fields: Human science, Letters, Economics, Law
- S fields: Engineering science

 ISBN: 978-1-61208-390-2

- Member composition: 2 H students and 2 S students Equations

Our experimental procedure is shown below.

(I) Interdisciplinary discussion for consensus building (40 minutes)

(II) Recognizing specialization using the video of (I)

This study obtains turn-taking information in (I) because turn-taking occurs whenever speakers engage in dialog. Specifically, we record the order of turn shift in the discussion and obtain turn-taking information. Information about awareness of specialization is acquired in (II). This study gathers information about awareness of specialization from each participant's decision. If more than one participant judged that the utterance was professional, this study treats it as having specialization. The situation of the experiment is shown in Figure 1.

D. Analysis

This study considers the rates at which specialization became apparently just after specialization pattern and proximate patterns. We define the rate as the "information giving rate." We can get information giving rate "P" as in the following expression when we define the number of specializations which participants recognized as Ra and the number of specializations which participants recognized and located just after specialization pattern or proximate patterns as Rs.

$$P = R_s / R_a \,. \qquad (2)$$

This study proposes two types of patterns to help participants recognize the specialization. Meanwhile, we can propose patterns that have high specialization rates. To compare proposed patterns and such patterns, we derive specialization rate for each pattern. We get the information about how many times each pattern appeared and participants recognized specialization after each pattern from our result. After that, we calculate the specialization rate of each pattern forms this information. We define such patterns as "higher patterns" and compare their information giving rate to analyze each pattern's feature.

The number of higher patterns is equal to the sum total of specialization pattern and proximate patterns. The number of utterance composing patterns is three or four. If the number of utterances is too few, the feature of interaction may be lost. On the other hand, if the number of utterances are too many, it may be difficult for participants to comprehend all of the patterns.

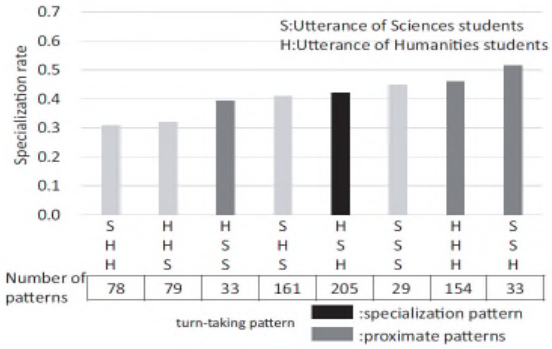


Figure 2. Specialization rate of each pattern (three utterances)

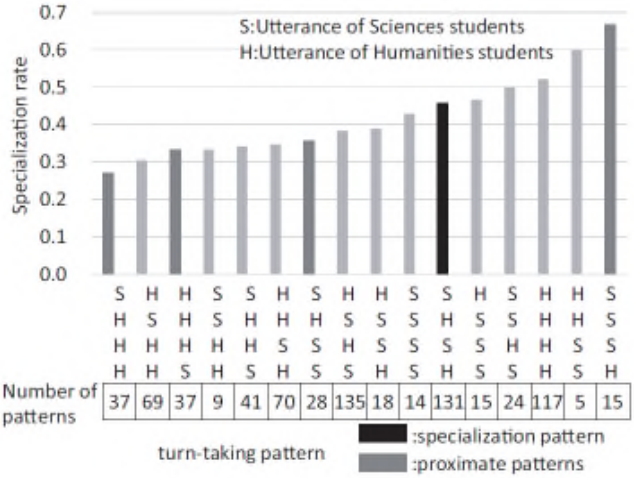


Figure 3. Specialization rate of each pattern (four utterances)

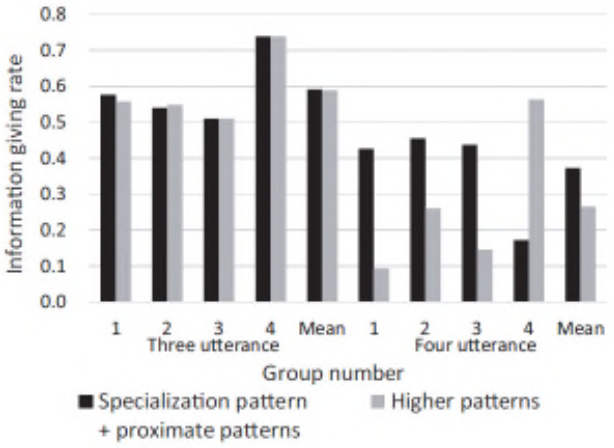


Figure 4. Information giving rate of each group

E. Result

Figure 2 shows specialization rates per patterns that are composed of three utterances. For example, item SSH shows the utterance just after the pattern contains the specialization at the rate of 50%. The number of each pattern's appearance is shown as "Number of patterns." Specialization pattern is indicated by a black bar and proximate patterns are depicted by gray bars. As a result, the utterance just after the specialization pattern contained specialization at the rate of nearly 40%. Some proximate patterns matched patterns whose specialization rates are higher.

Figure 3 shows specialization rates per patterns that are composed of four utterances. The graph composition is the same as in Figure 2. As shown in Figure 3, the utterance just after the specialization pattern contains the specialization at the rate of nearly 45%. One proximate pattern matched the pattern whose specialization rate is highest; however, the others show lower specialization rate patterns.

Figure 4 shows the information giving rates. Black bars show the rate at the time of presenting specialization pattern and proximate patterns and gray bars show the rates at the time of presenting higher patterns. As shown in Figure 4, in the case of three utterances, there is almost no difference between presenting specialization pattern and proximate patterns and presenting higher patterns. In the case of four utterances, Figure 4 shows presenting specialization pattern and proximate patterns can produce more information than presenting higher patterns except for group 4.

F. Considerlation

Figures 2 and 3 show there is some possibility that the proposed patterns can present good information from the perspective of the ease of recognition for every pattern because some specialization pattern and proximate patterns matched higher patterns. It appears that the reason some proximate patterns' specialization rates were low is the feature of specialization rates. Specialization rates become high not only when awareness of specialization transpired often but also when the patterns seldom appeared. This problem can be solved by providing more data for every pattern.

Figure 4 shows that information giving rates are influenced by the number of utterances that composes patterns. This result shows the possibility that we can change the amount of the information provided. It is necessary to provide a suitable quantity of information in accordance with the students' skill level. If the quantity of information is too high, participants may be confused. On the other hand, if the information is too scarce, the nature of the support system may be lost. We have confirmed that the proposed method can flexibly educate participants by changing the number of utterances that compose the relevant composing patterns. From this result, we can use the method proposing specialization pattern and proximate patterns for education support using awareness of specialization.

Moreover, Figure 4 shows specialization patterns and proximate patterns can transmit more information except in the case of group 2 of three utterances and group 4 of four utterances. This result shows that high specialization rate patterns cannot always give a great deal of information. One of the reasons for the result is same as the reason that proposed patterns did not match higher patterns. In other words, the cause is features of specialization rate. As in the other case, higher patterns that have been derived from all groups' data do not always match the patterns that have been derived from each group's data. This result shows that the proposed patterns can produce highly adaptable information.

In summary, these considerations show that if we can gather appropriate data, the proposed patterns can produce patterns that have a high specialization rate. Moreover, the proposed patterns can become a good support method under the same conditions. On the other hand, when the number of utterances which composes patterns is increased, proximate patterns' specialization rate and information giving rate decrease. This result shows that if we focus on more utterances, the number of patterns increases and awareness of specialization may diverge accordingly. In consequence, each pattern's specialization rate and information giving rate may be reduced. This phenomenon would occur when the number of fields is increased because this result is linked to an increase in the number of patterns. This study focuses on helping participants unfamiliar with interdisciplinary discussion recognize the specialization; therefore, this study presupposes that the case in which there are too many patterns is not in effect.

Meanwhile, there is a question of whether support interface is really necessary because participants recognized specialization in this experiment. This study gathers information about turn-taking and awareness of specialization separately; however, discussion and awareness should be performed at the same time. Participants who are unfamiliar with interdisciplinary communication do not always perform these tasks simultaneously. Moreover, even if one participant recognized specialization, the others might not be able to recognize the specialization. From these viewpoints, a support interface is needed.

V. TOWARD INTERFACE DESIGN AND DEVELOPMENT

A. Proposing interface design

This study shows that specialization pattern and proximate patterns can indicate patterns that have a high specialization rate with some accuracy. We propose an interface to help participants recognize the specialization based on the experiment result. Specifically, we propose the interface that calls participants attention to recognize the

 ISBN: 978-1-61208-390-2

specialization when utterance that includes specialization appears. We describe a detailed method.

- Store the information about turn-taking and awareness of specialization that were gathered from this experiment and other discussions in a database.
- Derive specialization pattern by using stored data and information from the present discussion.
- Propose that the next utterance may have specialization to participants if the present pattern matches proposal patterns.

These methods give participants a stronger chance to recognize whether some utterances have specialization and make this awareness easier. Figure 5 shows our proposition for the design of such an interface.

Moreover, this interface can give information to not only participants but also facilitators. Participants can recognize the specialization from the information through the facilitators.

B. The subject towards interface development

We describe the subject to develop the interface. At first, there is a problem with the data limitation. This study gets information from the discussion composed of 2 H participants and 2 S participants; however, the number of participant and participants' field ratio is not always the same. This interface can be applied to the case when these elements change though we need information about turn-taking and awareness of specialization from the other elements of discussions. We need more experiments to expand the useful range of our interface.

Second, there is a problem of how we determine the number of utterances that compose a specialization pattern. Figure 4 shows each case of information giving rate; however, the standard number of utterances that should be used remains indeterminate. This study focuses on three and four utterances; however, it is possible to increase the number of utterances unless the number of patterns increases too much. Therefore, it is necessary to consider the relationship between the standard of participants' skill level and the number of utterances that composes patterns to determine which information to give them.

C. Future works and subjects

The goal of this study is to develop an interface that is able to give all requisite information about awareness of specialization and help participants to be aware of and recognize it more readily. It seems that participants can get knowledge about the diversity of other participants' way of thinking and field and learn interdisciplinary communication through awareness of specialization by using such a system.

There are some obstacles to achieving this goal at present.

(i) Not all participants can recognize the specialization when information is presented.

(ii) Fields' classification and definition are very limited.

(iii) Classification is not applicable to all discussions. This section considers how to solve these problems. First, we focus on (i). This study leaves the decision of whether utterances have specialization to participants; no matter how well developed the system becomes, not all participant will be able to recognize the specialization when information is given. It is possible to make participants recognize the specialization by analyzing the contents of utterances to show more exactly timing. However, if this system provides more detailed information, there is a possibility that

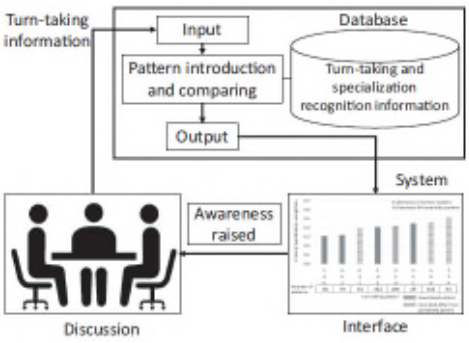


Figure 5. Interface image

participants cannot learn spontaneous awareness. It is necessary to consider how far we should provide support.

Second, we focus on (ii). As mentioned in Section III-A, the standards of fields' classification are based on Japanese universities' departments; therefore, it is unclear whether the proposed method can perform in foreign countries' educational environments. Moreover, this study presumes participants have only one specialization; however many experts have more than one specialization. It is necessary to examine the effect of this method based on field classifications considered common in the pertinent country and consider what happens when participants have multi specialization. We forecast that the effect of this method does not change when the fields' classification changes unless the number of classification is increase. In addition, it is expected that multi-specialized experts recognize more specialization than other participants and their specialization is easy to recognize.

Finally, we focus on (iii). Not all interdisciplinary discussion can separated in H and S. For example, discussions between Engineering experts and medical experts is considered as interdisciplinary discussion; however, these fields both belong S in this study's classification of fields. In that case, this method is not able to support participants; therefore, it is necessary to confirm whether the method can support participants in other systems of classifying fields. Moreover, if this method will be usable, we need new training data with a new classification to derive the proposed pattern.

 ISBN: 978-1-61208-390-2

VI. CONCLUSION

This study was aimed to support awareness of specialization in group discussion for learning interdisciplinary communication. Toward the goal, we proposed providing information based on transition of speakers' specialization following turn-taking as "pattern". We focused on when the pattern by which participants tend to recognize the specialization appears. We defined such patterns as "specialization pattern" and "proximate patterns" and examined whether these patterns can help participants to recognize the specialization. The result showed that, if we choose an appropriate number of utterances and have suitable data, specialization pattern and proximate patterns can help participants recognize the specialization in interdisciplinary discussion. We proposed an interface that indicates when specialization might appear in the next utterance when specialization pattern or proximate patterns appeared.

The proposed interface can be adapted to specific groups at the present. Further discussion is needed to gather information about turn-taking and awareness of specialization. Moreover, it is necessary to develop the interface to examine the effect of interdisciplinary communication learning through recognition of specialization.

ACKNOWLEDGMENT

We want to express special thanks to the students of the Osaka University who participated in our experiment.

REFERENCES

[1] D. H. Sonnenwald, "Communication roles that support collaboration during the design process," Design Studies, vol. 17, 1996, pp. 277– 301, ISSN: 0142-694X.

[2] P. Hall and L. Weaver, "Interdisciplinary education and teamwork: a long and winding road. Medical education," Journal of Medical Education, vol. 35, 2001, pp. 867–875, ISSN: 0308-0110.

[3] E. Yagi, "Toward Fusion between Face-to-face & Text-based Communication: Application to a Seminar on Science & Technology Communication," The Institute of Electronics, Information and Communication Engineers Technical Report, vol. 106, 2006, pp. 33–36, ISSN: 09135685.

[4] J. Mason and M. Spence, "Beyond mere knowledge of mathematics: The importance of knowing-to act in the moment," Educational Studies in Mathematics, vol. 38, 1999, pp. 135–161, ISSN: 0013-1954.

[5] Y. Fujigaki, "Difficulties in Interdisciplinary Research and Integration of Knowledge," Journal of Science policy and research management, vol. 10, 1996, pp. 73–83, ISSN: 09147020.

[6] K. Sumi and T. Nishida, "Communication Support System for User's Background Knowledge and the Context," The Transactions of the Institute of Electronics Information and Communication Engineers, vol. 84, 2001, pp. 1211–1221, ISSN: 09151915.

[7] H. Scholten, A. Kassahun, J. C. Refsgaard, T. Kargas, C. Gavardinas, and A. J. Beulens, "A methodology to support multidisciplinary model based water management," Environmental Modelling & Software, vol. 22, 2007, pp. 743–759, ISSN: 1364-8152.

[8] L. J. Miller, R. Gazan, and S. Still, "Unsupervised classification and visualization of unstructured text for the support of interdisciplinary collaboration," in Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing February 15–19, 2013, Baltimore, United States of America. ACM, Feb. 2014, pp. 1033–1042, ISBN: 978-14-50-32-54-00, doi: 10.1145/2531602.2531666, URL: http://dl.acm.org/citation.cfm?id=2531666 [accessed: 2014-10-27].

[9] H. Sacks, E. A. Schegloff, and G. Jefferson, "A simplest systematics for the organization of turn-taking for conversation," Language, vol. 50, 1974, pp. 696–735, ISSN: 00978507.

[10] H. Cao, O. Gapenne, and D. Aubert, "Accelerative effect of tactile feedback on turn-taking control in remote verbal communication," in Proceedings of CHI'13 Extended Abstracts on Human Factors in Computing Systems April 27– May 2, 2013, Paris, France. ACM, Apr. 2013, pp. 1033–1042, , ISBN: 978-1-45-03-19-52-2 , doi: 10.1145/2468356.2468639, URL: http://dl.acm.org/citation.cfm?id=2531666 [accessed: 2014-10-28].

[11] J. M. DiMicco, K. J. Hollenbach, and W. Bender, "Using Visualizations to Review a Group's Interaction Dynamics," in Proceedings of CHI '06 Extended Abstracts on Human Factors in Computing Systems April 24– 27, 2006, Montreal, Quebec, Canada. ACM, Apr. 2006, pp. 706–711, , ISBN: 1-59593-298-4 , doi: 10.1145/1125451.1125594, URL: http://dl.acm.org/citation.cfm?id=1125594 [accessed: 2014-10-28].

[12] N. Bouno and K. Takanashi, Eds., How to analysis a lot of people interaction. Omusya, Sep. 2009, ISBN: 978-4274207327.

[13] E. A. Schegloff, "Overlapping talk and the organization of turn-taking for conversation," Language in Society, vol. 29, 2000, pp. 1–63, ISSN: 00474045.

[14] K. Jisun, "An overview of turn-taking research: Some issues of "turn" and "turn-taking"(Part 4 Conversation research and Japanese language education)," Japanese Language Education, vol. 2002, 2002, pp. 205– 221, ISSN: 09174206.

[15] M. Umeki, "The educational separation history of Humanities and Science in Japan," Japanese Education Research Association, vol. 54, 1995, pp. 206–207.

[16] Y. Wajima, Y. Washida, and H. Ueda, "The difference in disciplines between humanities and science courses at a university affects the way of creative thinking," The Institute of Electronics, Information and Communication Engineers Technical Report, vol. 114, 2014, pp. 277– 282, ISSN: 0913-5685.

 ISBN: 978-1-61208-390-2

Recurrent Fuzzy Neural Network Controller Design for Ultrasonic Motor Rotor Angle Control

Tien-Chi Chen
Dept. of Computer and Communication
Kun Shan University
Tainan, Taiwan
e-mail: tchichen@mail.ksu.edu.tw

Tsai-Jiun Ren
Dept. of Information Engineering
Kun Shan University
Tainan, Taiwan
e-mail: cyrusren@mail.ksu.edu.tw

Yi-Wei Lou
Dept. of Electrical Engineering
Kun Shan University
Tainan, Taiwan
e-mail: oalatraszi@msn.com

***Abstract*—The ultrasonic motor (USM) has significant high precision, fast dynamic, simple structure and no electromagnetic interference features that are useful in many industrial, medical, robotic and automotive applications. The USM, however has nonlinear characteristics and dead-zone problems due to increasing temperature and motor drive issues under various operating conditions. To overcome these problems a recurrent fuzzy neural network controller (RFNNC) combined with a compensated controller with adjustable parameters and on line learning algorithm is presented in this paper. The proposed control scheme can take the nonlinearity into account and compensate for the USM dead-zone. The proposed control scheme provides robust performance against parameter variations. The experimental results demonstrate the effectiveness of the proposed USM control scheme.**

Keywords- ultrasonic motor; recurrent fuzzy neural network; compensated controller; adjustable parameters; on line learning algorithm.

I. INTRODUCTION

The USM has many excellent features such as high precision, fast dynamics, simple structure, compactness in size and no electromagnetic interference, which is useful in many industrial, medical and automotive applications [1-4]. However, the USM has nonlinear characteristics and a dead-zone problem due to the large static friction torque appearing at low speed [5]. Hence, it is difficult to design a perfect controller to accurately control the USM at all times.

Conventional PI controllers for common motors have the advantages of simple and easy design, high-stability margin and high-reliability when the controllers are tuned properly [6-7]. However, the PI controller cannot maintain these virtues at all times. The USM has nonlinear speed characteristics which vary with the drive operating conditions. A dynamic controller with adjustable parameters and on line learning algorithms is therefore suggested for nonlinear or uncertain dynamics systems [8-11].

This paper presents a new control scheme for USM rotor angle control, the RFNNC combined with a compensated controller that has adjustable parameters and an on-line learning algorithm to overcome the nonlinear characteristic and dead-zone problem. The RFNNC is effective in handling nonlinear motor characteristic variations due to connecting weight updating in the RFNNC. Furthermore, a compensated controller is presented to compensate the USM dead-zone effect, which deteriorates the dynamic response. The proposed control scheme can take the nonlinearity into account and compensate for the USM dead-zone. Robust performance against parameter variations is obtained by the proposed approach. The usefulness and validity of the proposed control scheme is examined through experimental results. The experimental results reveal that the proposed control scheme maintains good, stable performance under different motion conditions.

II. THE PROPOSED CONTROL SCHEME

Fig. 1 shows the proposed control scheme for the USM combined with the RFNNC and compensated controller that has adjustable parameters and an on-line learning algorithm to overcome the nonlinear characteristic and dead-zone problem, where θ_c the USM rotor angle command, θ_r the USM rotor angle, u_R the RFNNC output control force, u_C the compensated controller output control force and u is the total control force.

A. RFNNC Structure

Fig. 2 shows the RFNNC structure, comprised of an input layer, membership layer, rule layer and output layer.

The RFNNC mathematical model is summarized as follows.

(1) Input layer

The RFNNC inputs are $e = \theta_c - \theta_r$ and $\dot{e}$. The outputs $x_{e,i}^1$ and $x_{\dot{e},i}^1$ can be expressed as:

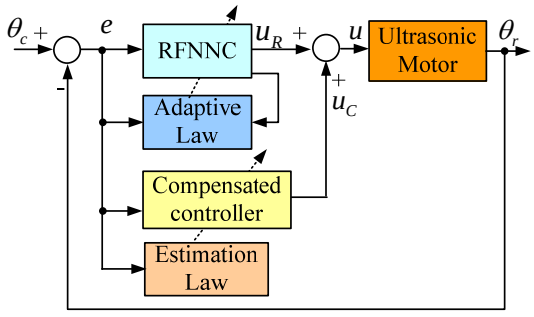


Figure 1. The proposed control scheme for the USM.

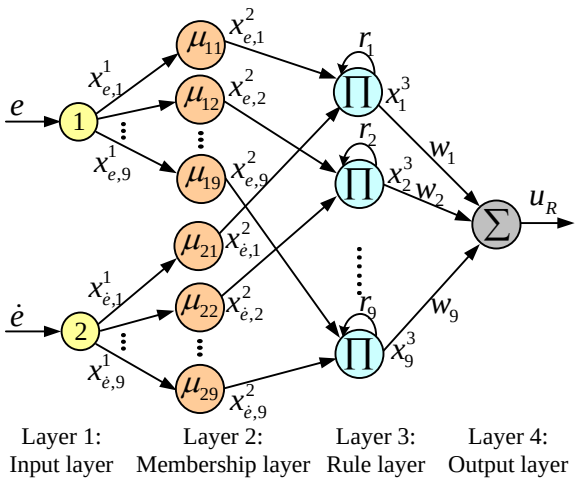


Figure 2. The RFNNC structure.

$$x_{e,i}^1 = e, \quad i = 1,...,9 \tag{1}$$

$$x_{\dot{e},i}^1 = \dot{e}, \quad i = 1,...,9 \tag{2}$$

(2) Membership layer

The membership layer outputs $y_{e,i}^2$ and $y_{\dot{e},i}^2$ can be expressed as a Gaussian function:

$$x_{e,i}^2 = \exp\left(-\left(\frac{x_{e,i}^1 - s_{e,i}}{z_{e,i}}\right)^2\right), \quad i = 1,...,9 \tag{3}$$

$$x_{\dot{e},i}^2 = \exp\left(-\left(\frac{x_{\dot{e},i}^1 - s_{\dot{e},i}}{z_{\dot{e},i}}\right)^2\right), \quad i = 1,...,9 \tag{4}$$

where $\mathbf{S} = \left[s_{e,1},...,s_{e,9},s_{\dot{e},1},...,s_{\dot{e},9}\right]^T$ the mean Gaussian function vector and $\mathbf{Z} = \left[z_{e,1},...,z_{e,9},z_{\dot{e},1},...,z_{\dot{e},9}\right]^T$ is the Gaussian function standard deviation vector.

(3) Rule layer

The rule layer outputs are expressed as:

$$x_i^3(t) = \left(1 + \frac{1}{1+\exp^{-r_i x_i^3(t-1)}}\right) x_{e,i}^2(t) x_{\dot{e},i}^2(t), \quad i = 1,...,9 \tag{5}$$

where $\mathbf{R} = \left[r_1,...,r_9\right]$ is the recurrent weight vector.

(4) Output layer

The output layer outputs u_R can be expressed as:

$$u_R = \sum_{i=1}^{9} w_i x_i^3 = \mathbf{W}^T \mathrm{T}(x,s,z,r) \tag{6}$$

where $\mathbf{T}(x,s,z,r) = \left[x_1^3,...,x_9^3\right]^T$ is fuzzy rule function vector and $\mathbf{W} = \left[w_1,...,w_9\right]^T$ is the adjustable output weight vector.

B. RFNNC and Compensated Controller Design

Consider the USM nonlinear dynamic system is expressed as:

$$\ddot{y} = f(y) + g(y)u(t) + d(t) \tag{7}$$

where $f(y)$ and $g(y)$ are unknown USM functions and assume they are bounded, $u(t)$ the control input, $d(t)$ the external disturbance.

The control goal is to design a RFNNC such that the USM rotor angle tracks the reference model output angle. The tracking error vector is first defined as

$$\mathbf{E} = \left[e,\ \dot{e}\right]^T \tag{8}$$

From (7) and (8), an ideal controller can be chosen as

$$u^*(t) = \frac{1}{g_n(y)}\left[\ddot{y} - f_n(y) - d_n(t) + \mathbf{K}^T\mathbf{E}\right] \tag{9}$$

In (9) , $\mathbf{K} = \left[k_2,\ k_1\right]^T$, in which k_1 and k_2 are positive constants. Applying (7) to (9), the error dynamics can be expressed as

$$\ddot{e} + k_1\dot{e} + k_2 e = 0 \tag{10}$$

If K is chosen to correspond to the coefficients of a Hurwitz polynomial, that is a polynomial whose roots lie strictly in the open left half of the complex plane, then the result achieved where $\lim_{t\to\infty} e(t) = 0$ for any initial conditions. Nevertheless, the functions $f(y)$ and $g(y)$ aren't accurate known and the external load disturbances are perturbed. Thus, the ideal controller $u^*(t)$ cannot be practical implemented. Therefore, the RFNNC will be designed to approximate this ideal controller.

The proposed control scheme for the USM combined by RFNNC and compensated controller is show in Fig. 1. The control force for the USM is the following form:

$$u = u_R + u_C \tag{11}$$

where the RFNNC output control force u_R is the main tracking control to approximate the ideal control force $u^*(t)$. From (7), (9) and (11), an error equation is rewritten as:

$$\dot{\mathbf{E}} = \mathbf{A}\mathbf{E} + \mathbf{B}(u^* - u_R - u_C) \tag{12}$$

where $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -k_2 & -k_1 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 0 \\ g(y) \end{bmatrix}$.

Assume that an optimal RFNNC exists to approximate the ideal control force such that

$$u^* = u_R^*(e,\mathbf{W}^*,\mathbf{S}^*,\mathbf{Z}^*,\mathbf{R}^*) + \varepsilon = \mathbf{W}^{*T}\mathbf{T}^* + \varepsilon \tag{13}$$

where ε is a minimum reconstructed error, $\mathbf{W}^*,\mathbf{S}^*,\mathbf{Z}^*,\mathbf{R}^*$ and $\mathbf{T}^*$ are optimal parameters of $\mathbf{W},\mathbf{S},\mathbf{Z},\mathbf{R}$ and $\mathbf{T}$, respectively. Thus, the control force is assumed to take the following form:

$$u = u_R(e,\hat{\mathbf{W}},\hat{\mathbf{S}},\hat{\mathbf{Z}},\hat{\mathbf{R}}) + u_C = \hat{\mathbf{W}}^T\hat{\mathbf{T}} + u_C \tag{14}$$

 ISBN: 978-1-61208-390-2

where $\hat{\mathbf{W}}$, $\hat{\mathbf{S}}$, $\hat{\mathbf{Z}}$, $\hat{\mathbf{R}}$ and $\hat{\mathbf{T}}$ are the optimal parameter estimations provided by tuning the algorithms to be introduced later. Subtracting (14) from (13) an approximation error $\tilde{u}$ is obtained as

$$\begin{aligned}\tilde{u} &= u^* - u = \mathbf{W}^{*T}\mathbf{T}^* + \varepsilon - \hat{\mathbf{W}}^T\hat{\mathbf{T}} - u_C \\ &= \tilde{\mathbf{W}}^T\mathbf{T}^* + \hat{\mathbf{W}}^T\tilde{\mathbf{T}} + \varepsilon - u_C\end{aligned} \tag{15}$$

where $\tilde{\mathbf{W}} = \mathbf{W}^* - \hat{\mathbf{W}}$ and $\tilde{\mathbf{T}} = \mathbf{T}^* - \hat{\mathbf{T}}$. The linearization technique transforms the multidimensional receptive-field basis functions into a partially linear form such that the expansion of $\tilde{\mathbf{T}}$ in Taylor series becomes

$$\tilde{\mathbf{T}} = \left[\tilde{x}_1^3, ..., \tilde{x}_9^3\right]^T = \mathbf{T}_s\tilde{\mathbf{S}} + \mathbf{T}_z\tilde{\mathbf{Z}} + \mathbf{T}_r\tilde{\mathbf{R}} + \mathbf{O}_v \tag{16}$$

where $\tilde{x}_i^3 = x_i^{3*} - \hat{x}_i^3$, x_i^{3*} is the optimal parameter, $\hat{x}_i^3$ is the estimated parameter of x_i^{3*}, $\tilde{\mathbf{S}} = \mathbf{S}^* - \hat{\mathbf{S}}$, $\tilde{\mathbf{Z}} = \mathbf{Z}^* - \hat{\mathbf{Z}}$, $\tilde{\mathbf{R}} = \mathbf{R}^* - \hat{\mathbf{R}}$, $\mathbf{O}_v$ is higher-order terms, $\mathbf{T}_s = \left[\partial x_1^3/\partial s, ..., \partial x_9^3/\partial s\right]_{s=\hat{s}}^T$, $\mathbf{T}_z = \left[\partial x_1^3/\partial z, ..., \partial x_9^3/\partial z\right]_{z=\hat{z}}^T$, $\mathbf{T}_r = \left[\partial y_1^3/\partial r, ..., \partial y_k^3/\partial r\right]_{r=\hat{r}}^T$.

(16) can be rewritten as

$$\mathbf{T}^* = \hat{\mathbf{T}} + \mathbf{T}_s\tilde{\mathbf{S}} + \mathbf{T}_z\tilde{\mathbf{Z}} + \mathbf{T}_r\tilde{\mathbf{R}} + \mathbf{O}_v \tag{17}$$

Substituting (17) into (15), it can be rewritten as:

$$\tilde{u} = \tilde{\mathbf{W}}^T\hat{\mathbf{T}} + \hat{\mathbf{W}}^T(\mathbf{T}_s\tilde{\mathbf{S}} + \mathbf{T}_z\tilde{\mathbf{Z}} + \mathbf{T}_r\tilde{\mathbf{R}}) - u_C + D \tag{18}$$

where $D = \tilde{\mathbf{W}}^T(\mathbf{T}_s\tilde{\mathbf{S}} + \mathbf{T}_z\tilde{\mathbf{Z}} + \mathbf{T}_r\tilde{\mathbf{R}}) + \mathbf{W}^{*T}\mathbf{O}_v + \varepsilon$ is the uncertainty term and this term is assumed to be bounded with a small positive constant β （let $|D| \le \beta$）. From (15) and (18), (12) can be rewritten as

$$\begin{aligned}\dot{\mathbf{E}} &= \mathbf{AE} + \mathbf{B}(u^* - u) = \mathbf{AE} + \mathbf{B}\tilde{u} \\ &= \mathbf{AE} + \mathbf{B}\left[\tilde{\mathbf{W}}^T\hat{\mathbf{T}} + \mathbf{W}\hat{\mathbf{W}}^T(\mathbf{T}_s\tilde{\mathbf{S}} + \mathbf{T}_z\tilde{\mathbf{Z}} + \mathbf{T}_r\tilde{\mathbf{R}}) - u_C + D\right]\end{aligned} \tag{19}$$

Consider the USM dynamic system is represented by (7) if the RFNNC control law is designed as (14) with the adaptive laws for networks parameters, as shown in (20)–(23), and the compensated controller control law is designed as (24) with the estimation law given in (25). The stability of the proposed control scheme can then be guaranteed, where η_1, η_2, η_3, η_4, η_5 are strictly positive constants.

$$\dot{\hat{\mathbf{W}}} = \eta_1\hat{\mathbf{T}}\mathbf{E}^T\mathbf{PB} \tag{20}$$

$$\dot{\hat{\mathbf{S}}} = \eta_2\mathbf{T}_s^T\hat{\mathbf{W}}\mathbf{E}^T\mathbf{PB} \tag{21}$$

$$\dot{\hat{\mathbf{Z}}} = \eta_3\mathbf{T}_z^T\hat{\mathbf{W}}\mathbf{E}^T\mathbf{PB} \tag{22}$$

$$\dot{\hat{\mathbf{R}}} = \eta_4\mathbf{T}_r^T\hat{\mathbf{W}}\mathbf{E}^T\mathbf{PB} \tag{23}$$

$$u_C = \hat{\beta}\,\mathrm{sgn}(\mathbf{E}^T\mathbf{PB}) \tag{24}$$

$$\dot{\hat{\beta}} = \eta_5\left|\mathbf{E}^T\mathbf{PB}\right| \tag{25}$$

Proof: Define a Lyapunov function candidate as

$$\begin{aligned}V(t) = &\frac{1}{2}\mathbf{E}^T\mathbf{PE} + \frac{1}{2\eta_1}tr(\tilde{\mathbf{W}}^T\tilde{\mathbf{W}}) + \frac{1}{2\eta_2}\tilde{\mathbf{S}}^T\tilde{\mathbf{S}} \\ &+ \frac{1}{2\eta_3}\tilde{\mathbf{Z}}^T\tilde{\mathbf{Z}} + \frac{1}{2\eta_4}\tilde{\mathbf{R}}^T\tilde{\mathbf{R}} + \frac{1}{2\eta_5}\tilde{\beta}^2\end{aligned} \tag{26}$$

where $\mathbf{P}$ is a symmetric positive definite matrix which satisfies the following Lyapunov equation

$$\mathbf{A}^T\mathbf{P} + \mathbf{PA} = -\mathbf{Q} \tag{27}$$

where $\mathbf{Q}$ is a positive definite matrix. Here, the uncertainty boundary estimation error is defined as $\tilde{\beta} = \beta - \hat{\beta}$. Taking the Lyapunov function (26) differential and using (18) and (27), it is concluded that

$$\begin{aligned}\dot{V}(t) = &-\frac{1}{2}\mathbf{E}^T\mathbf{QE} + \frac{1}{2}(\mathbf{E}^T\mathbf{PB} + \mathbf{B}^T\mathbf{PE})\tilde{u} - \frac{1}{\eta_1}\tilde{\mathbf{W}}^T\dot{\hat{\mathbf{W}}} \\ &- \frac{1}{\eta_2}\dot{\hat{\mathbf{S}}}^T\tilde{\mathbf{S}} - \frac{1}{\eta_3}\dot{\hat{\mathbf{Z}}}^T\tilde{\mathbf{Z}} - \frac{1}{\eta_4}\dot{\hat{\mathbf{R}}}^T\tilde{\mathbf{R}} - \frac{1}{\eta_5}\tilde{\beta}\dot{\hat{\beta}} \\ = &-\frac{1}{2}\mathbf{E}^T\mathbf{QE} + \mathbf{E}^T\mathbf{PB}\left[\tilde{\mathbf{W}}^T\hat{\mathbf{T}} + \hat{\mathbf{W}}^T(\mathbf{T}_s\tilde{\mathbf{S}} + \mathbf{T}_z\tilde{\mathbf{Z}} + \mathbf{T}_r\tilde{\mathbf{R}}) - u_C + D\right] \\ &- \frac{1}{\eta_1}\tilde{\mathbf{W}}^T\dot{\hat{\mathbf{W}}} - \frac{1}{\eta_2}\dot{\hat{\mathbf{S}}}^T\tilde{\mathbf{S}} - \frac{1}{\eta_3}\dot{\hat{\mathbf{Z}}}^T\tilde{\mathbf{Z}} - \frac{1}{\eta_4}\dot{\hat{\mathbf{R}}}^T\tilde{\mathbf{R}} - \frac{1}{\eta_5}\tilde{\beta}\dot{\hat{\beta}}\end{aligned} \tag{28}$$

Take (20)-(25) into (28), the derivative of V can be rewritten as

$$\begin{aligned}\dot{V}(t) &= -\frac{1}{2}\mathbf{E}^T\mathbf{QE} + \mathbf{E}^T\mathbf{PB}D - \mathbf{E}^T\mathbf{PB}u_C - \frac{1}{\eta_5}(\beta - \hat{\beta})\dot{\hat{\beta}} \\ &\le -\frac{1}{2}\mathbf{E}^T\mathbf{QE} + \mathbf{E}^T\mathbf{PB}D - \hat{\beta}\left|\mathbf{E}^T\mathbf{PB}\right| - (\beta - \hat{\beta})\left|\mathbf{E}^T\mathbf{PB}\right| \\ &\le -\frac{1}{2}\mathbf{E}^T\mathbf{QE} - \left|\mathbf{E}^T\mathbf{PB}\right|(\beta - |D|) \le 0\end{aligned} \tag{29}$$

This implies that $\mathbf{E}(t)$ will converge to zero as $t \to \infty$. As a result the stability of the proposed control system can be guaranteed.

III. EXPERIMENTAL RESULTS

The USM experimental setup is shown in Fig. 3. The TMS320F2812 digital signal processor (DSP) is used for design the proposed control scheme. The USM driver is used to drive the USM, which encoder signal of rotor angle and speed are fed back to the TMS320F2812 to construct closed-loop control.

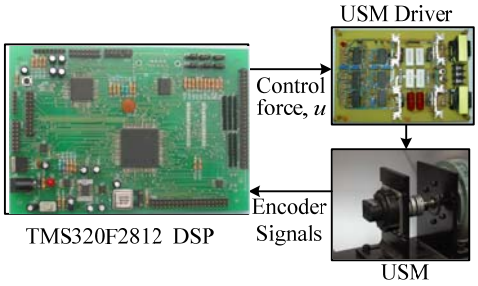


Figure 3. The USM experiment setup.

A. A square angle command from -45 to 45 degree

In Figs. 4, 5 and 6, the USM rotor angle command is a periodic square from -45 to 45 degree. The USM rotor angle and speed responses using the proposed control scheme are depicted in Fig. 4(a). The angle error between the angle command and USM rotor angle is depicted in Fig. 4(b). By observing the experimental result in Fig. 4, the tracking errors can both converge to an acceptable region and the control performance is excellent. The proposed controller retains control performance and has no dead-zone in the construction.

The USM rotor angle and speed responses using the RFNNC without compensated controller are depicted in Fig. 5(a). The angle error is depicted in Fig. 5(b). Compared with the RFNNC without compensated controller in Fig. 5, it can be seen that the RFNNC without compensated controller can achieve barely satisfactory response in each state. But, it exhibits a chattering phenomenon due to the dead-zone especially in slow speed nearby zero.

The USM rotor angle and speed responses using the PI control are depicted in Fig. 6(a). The angle error is depicted in Fig. 6(b). In Fig. 6 illustrated that the PI control has a chattering phenomenon that cannot maintain the control performance when the parameter plant variation is large and the drive conditions change frequently.

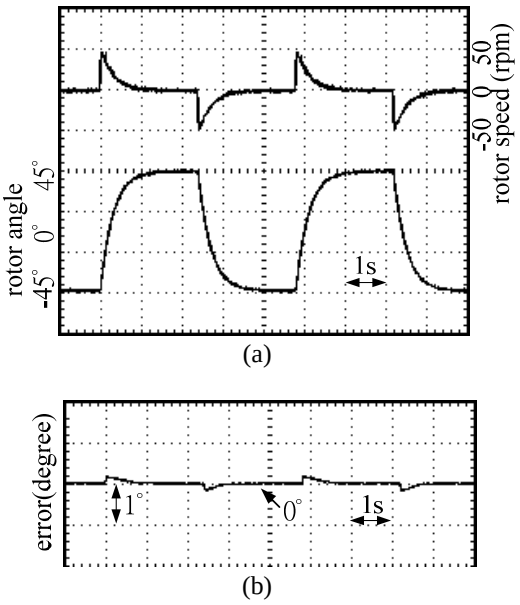


Figure 4. The experimental results using the proposed control scheme for a periodic square command from -45 to 45 degree. (a)USM rotor angle and speed responses, (b) angle error.

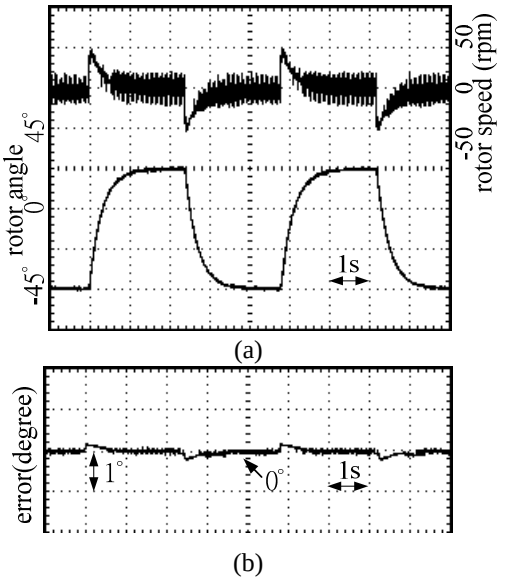


Figure 5. The experimental results using the RFNNC without compensated controller for a periodic square command from -45 to 45 degree. (a) USM rotor angle and speed responses, (b) angle error.

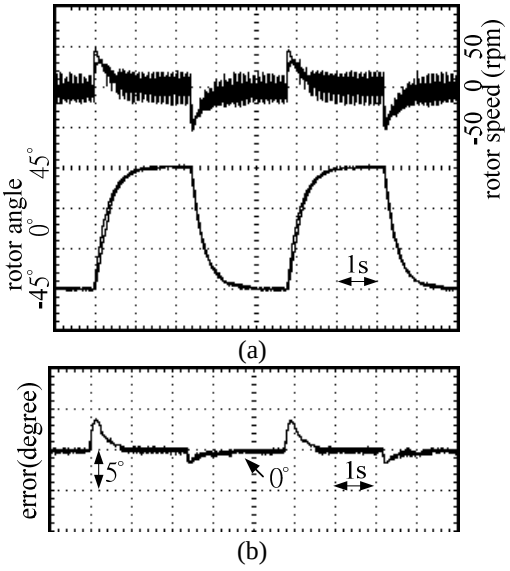


Figure 6. The experimental results using the PI control for a periodic square command from -45 to 45 degree. (a) USM rotor angle and speed responses, (b) angle error.

B. A sinusoidal angle command from -45 to 45 degree

In Figs. 7, 8 and 9 show the USM rotor angle command is a sinusoidal from -45 to 45 degrees. The USM rotor angle and speed responses using the proposed control scheme are depicted in Fig. 7(a). The angle error between angle command and USM rotor angle is depicted in Fig. 7(b). Fig. 7 shows the experimental results with the proposed control scheme that the favorable tracking error is quickly reduced to zero.

 ISBN: 978-1-61208-390-2

The USM rotor angle and speed responses using the RFNNC without compensated controller are depicted in Fig. 8(a). The angle error is depicted in Fig. 8(b). Compared with the RFNNC without compensated controller in Fig. 8, it can be seen that the RFNNC without compensated controller can achieve barely satisfactory response which is similar to the proposed control scheme. However, it exhibits a chattering phenomenon due to the dead-zone especially in slow speed nearby zero.

The USM rotor angle and speed responses using the PI control are depicted in Fig. 9(a). The angle error is depicted in Fig. 9(b). Fig. 9 illustrates that the PI controller tracking response is slower and the steady state error is larger than that of proposed control schemes. The drawbacks of the PI control are interference with the dead-zone and the motor speed has a serious chattering phenomenon at slow speed near zero.

C. A constant speed command of 100 rpm with 0.5 N-m load

Fig. 10 shows the USM rotor speed command is constant at 100 rpm with 0.5N-m load. Fig. 10(a) shows the 0.5N-m load applied to the USM rotor. The USM rotor speed and speed error between the speed command and USM rotor speed using the proposed control scheme are depicted in Fig. 10(b). The USM rotor speed and speed error between the speed command and USM rotor speed using the PI control are depicted in Fig. 10(c). The experimental results in Fig.10 depict that the PI control has larger speed ripple response than the proposed control scheme. Based on the experimental results, the control performance of the proposed control scheme is better than that of PI control.

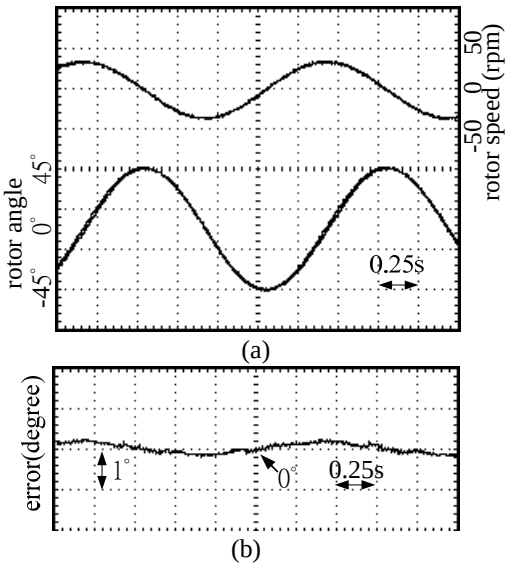


Figure 7. The experimental results using the proposed control scheme for a sinusoidal command from -45 to 45 degree. (a)USM rotor angle and speed responses, (b) angle error.

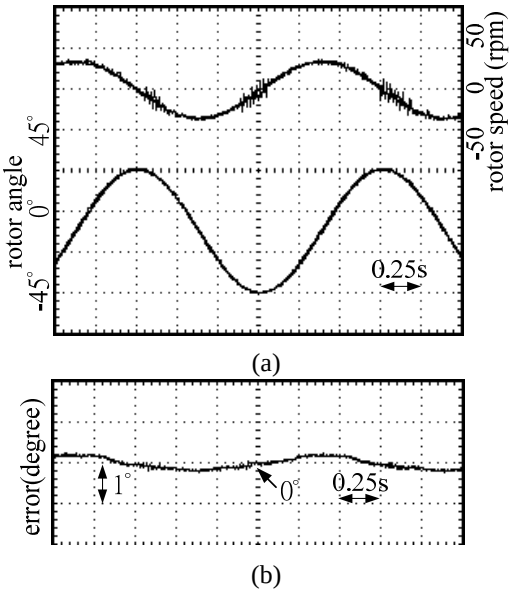


Figure 8. The experimental results using the RFNNC without compensated controller for a sinusoidal command from -45 to 45 degree. (a) USM rotor angle and speed responses, (b) angle error.

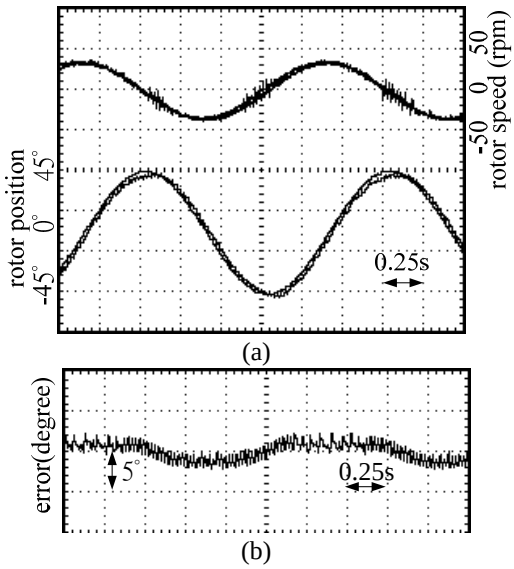


Figure 9. The experimental results using the PI control for a sinusoidal command from -45 to 45 degree. (a) USM rotor angle and speed responses, (b) angle error.

 ISBN: 978-1-61208-390-2

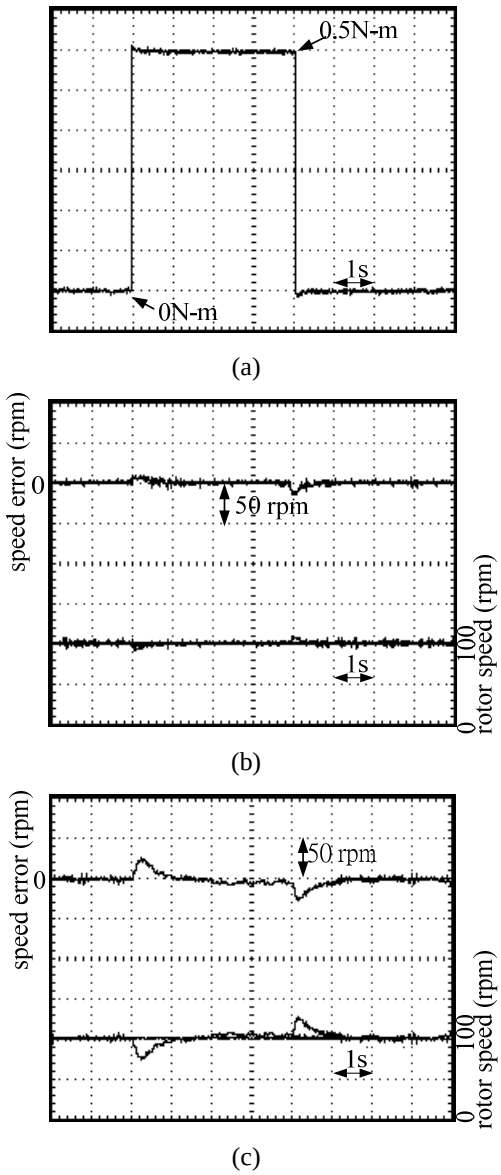


Figure 10. The experimental results for the USM rotor speed command of 100 rpm with 0.5N-m load. (a) applied load, (b) rotor speed and speed error using the proposed control scheme, (c) rotor speed and speed error using the PI control.

IV. CONCLUSIONS

A novel control scheme, using RFNNC combined with a compensated controller, has been applied to USM rotor angle control. The RFNNC is designed in handling nonlinear motor characteristic variations to track the reference angle. A compensated controller is designed to compensate the USM dead-zone effect, which deteriorates the dynamic response. The RFNNC parameters are tuned in the Lyapunov sense; thus, the system stability can be guaranteed. Experimental results show that the proposed control scheme is better than that of PI control.

ACKNOWLEDGMENT

The authors would like to express their appreciation to Ministry of Science Technology for supporting under contact MOST 103-2627-E-168 -001.

REFERENCES

[1] M. A. Tavallaei, Y. Thakur, S. Haider, and M. Drangova," Magnetic-resonance-imaging-compatible remote catheter navigation system," IEEE Trans. Biomedical Engineering, vol. 60, no. 4, pp. 899-905, April 2013.

[2] M. Guo, J. Hu, H. Zhu, C. Zhao, and S. Dong, "Three-degree-of-freedom ultrasonic motor using a 5-mm-diameter piezoelectric ceramic tube," IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency control, vol. 60, no. 7, pp. 1446-1452, July 2013.

[3] H. Y. Wang, K. C. Fan, J. K. Ye, and C. H. Lin, "A long-stroke nanopositioning control system of the coplanar stage," IEEE Trans. Mechatronics, vol. 19, no. 1, pp. 348-356, February 2014.

[4] P. Ci, G. Liu, Z. Chen, S. Zhang, and S. Dong, "High-order face-shear modes of elaxor-PbTiO3 crystals for poezoelectric motor applications," Applied Physics Letters, vol. 104, no. 24, pp. 242911 - 242911-4, July 2014

[5] F. Giraud, P. Sandulescu, M. Amberg, B. Lemaire-Semail, and F. Ionescu, "Modeling and compensation of the internal friction torque of a travelling wave ultrasonic motor," IEEE Trans. Haptics, vol. 4, no. 4, pp. 327-331, 2011.

[6] R. E. Precup, R. C. David, E. M. Petriu, M. B. Radac, and S. Preitl, "Adaptive GSA-based optimal tuning of PI controlled servo systems with reduced process parametric sensitivity, robust stability and controller robustness," IEEE Trans. Cybernetics, vol. 44, no. 11, pp. 1997-2009, November 2013.

[7] T. S. Franklin, J. J. F. Cerqueira, and E. S. de Santana, "Fuzzy and PI controllers in pumping water system using photovoltaic electric generation," IEEE Latin America Transactions, vol. 12, no. 6, pp. 1049-1054, September 2014.

[8] J. Jingzhuo, and B. Liu, "Optimum efficiency control of traveling-wave ultrasonic motor system," IEEE Trans. Industrial Electronics, vol. 58, no. 10, pp. 4822-4829, October 2011.

[9] N. T. Hieu, S. Odomari, T. Yoshida, T. Senjyu, and A. Yona, "Nonlinear adaptive control of ultrasonic motors considering dead-zone," IEEE Trans. Industrial Informatics, vol. 9, no. 4, pp. 1847-1854, November 2013.

[10] S. Zhou and Z. Yao , "Design and optimization of a modal-independent linearultrasonic motor," IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency control, vol. 61, no. 3, pp. 535-546, March 2014.

 ISBN: 978-1-61208-390-2

On the Generation of Privatized Synthetic Data Using Distance Transforms

Kato Mivule
Bowie State University
Bowie, MD, USA
kmivule@gmail.com

***Abstract*—Organizations have interest in research collaboration efforts that involve data sharing with peers. However, such partnerships often come with confidentiality risks that could involve insider attacks and untrustworthy collaborators who might leak sensitive information. To mitigate such data sharing vulnerabilities, entities share privatized data with retracted sensitive information. However, while such data sets might offer some assurances of privacy, maintaining the statistical traits of the original data, is often problematic, leading to poor data usability. Therefore, in this paper, a confidential synthetic data generation heuristic, that employs a combination of data privacy and distance transforms techniques, is presented. The heuristic is used for the generation of privatized numeric synthetic data, while preserving the statistical traits of the original data. Empirical results from applying unsupervised learning, using k-means, to test the usability of the privatized synthetic data set, are presented. Preliminary results from this implementation show that it might be possible to generate privatized synthetic data sets, with the same statistical morphological structure as the original, using data privacy and distance transforms methods**.

Keywords-Privatized synthetic data generation; Data privacy; Distance transforms; k-means clustering

I. INTRODUCTION

Research collaboration among organizations often involves the sharing of data, however, the issue of data confidentiality is often an impediment in such partnerships. To safely engage in joint research ventures, entities often retract sensitive and private information from the shared data, which reduces usability, despite confidentiality assurances. Yet still, another method used to address such data sharing vulnerabilities, is to generate privatized synthetic data sets that retain the statistical traits of the original data while at the same time ensuring privacy. In this paper, we present a confidential synthetic data generation heuristic, that employs data privacy and distance transforms methods, for the generation of privatized synthetic data while maintaining some of the statistical traits of the original data. In the initial stage, we apply distance transforms to extract the coefficients with the needed traits, from the original data (noisy data in this case), we then add the coefficients to a noisy data set, generating a privatized synthetic data set. Filtering is then applied to the privatized synthetic data set, to reduce noise and enhance usability. We then apply unsupervised learning, using k-means clustering, to test the usability of the privatized synthetic data set. We present preliminary results showing that it might be possible to generate privatized synthetic data sets, with the same statistical skeletal structure as the original, using distance transforms. Therefore, the main goal of this investigation is to employ data privacy, distance transforms, and k-means clustering approaches in the production of privatized synthetic data with similar statistical traits as the original data. The rest of the paper is organized as follows, in Section II, background and related work is given, while Section III talks about the methodology. In Section IV, a discussion of the experiment and results is done, and lastly, in Section V, the conclusion is given.

II. BACKGROUND AND RELATED WORK

Not much work exists on the application of distance transforms for privatized synthetic data generation. The technique of distance transforms has largely been used for applications in the image-processing domain. However, a look at works by researchers in the signal processing domain shows that techniques, such as, discrete cosine transforms, have been proposed for privacy preservation applications [12][13][14][15]. For instance, Mukherjee, Chen, and Gangopadhyay (2006) suggested using Fourier-related transforms, to enhance Euclidean distance-based algorithms for privacy preservation in data mining applications [1]. Of the privacy preservations problems that Mukherjee et al., (2006) observed, was that while data allocations of the original data could be maintained in the confidential data set, the space between each data point in the confidential data set is not kept, which often leads to diminished cluster outcomes [1]. Moreover, Mukherjee et al., (2006) noted that one of the benefits of using signal processing techniques, such as discrete cosine, is that the Euclidean distance among points in the confidential data set could be maintained, resulting in improved clustering results [1]. *Distance transforms:* Distance transforms, a skeletonization process, proposed by Rosenfeld and Pfaltz (1968), and mainly used in the image processing domain, is a morphological technique that alters a binary image made up of object O foreground and object O' background pixels into a resulting skeletal figure in which each object pixel has an analogous value to the smallest amount of space from the background object O'. Distance transforms can be expressed using the following formula [2][3][4]:

$$D(p) = \text{minimum}\{distance(p,q),\ q \in O\} \qquad (1)$$

The symbols O and O' represent the object foreground and background; $distance(p,q)$ represents the space

 ISBN: 978-1-61208-390-2

between pixel p and q . The least distance $minimum\{distance(p,q)\}$ is often required; and $D(p)$ symbolized the distance point at pixel p [3]. Euclidean distance is used for the morphological process [5][6][7].

III. METHODOLOGY

In this investigation, rather than apply discrete cosine transforms, as in [1], we apply distance transforms on a noisy data set with very similar traits to the original data. We use the following implementation phases in the generation of the privatized synthetic data as illustrated in Figure 1:

- *Phase 1:* In the first step of the process, a noisy data set, instead of the original, is generated so as to add an extra layer of privacy and make it difficult for reconstruction attacks. In case of a successful reconstruction attack, what the adversary gets is the noisy data, assuming no prior insider knowledge. Data privacy, in this first step, is achieved using noise addition [9], with a distribution $\varepsilon \sim N(\mu = 1, \sigma = 0.2)$ – generating a noisy data set with similar statistical traits to the original [10].

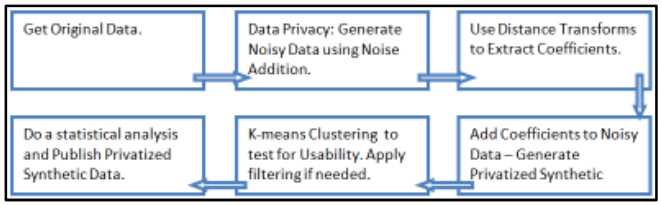


Figure 1: The Privatized Synthetic Data Generation Process

- *Phase 2:* In the second step, Distance transforms is applied on the noisy data set to extract coefficients.
- *Phase 3:* During the third step, the extracted coefficients are then added to back to the noisy data, generating the privatized synthetic data.
- *Phase 4:* In the fourth step, in order to reduce any excess noise, the moving average filtering is applied on the privatized synthetic data.
- *Phase 5:* In the fifth step of the process, we apply k-means clustering using Euclidean distance, to test the usability of the privatized synthetic data set, in comparison with the original data.
- *Phase 6:* In the final step, statistical analysis of both the original and privatized data sets is done, and the privatized synthetic data is published.

IV. RESULTS AND DISCUSSION

The Fisher Iris data set used in this experiment, comprised of 150 data points, five attributes, namely, sepal length, sepal width, petal length, petal width, and class attribute, with three classes, namely, Setosa, Versicolar, and Virginica [8]. The plots in Figure 2 illustrate series for the Sepal length, Sepal width, Petal length, and Petal width correspondingly, before and after application of distance transforms. For each plot, the upper series symbolizes the privatized synthetic Fisher-Iris data, the middle series symbolizes the noisy Fisher-Iris data used to generate the privatized synthetic data, and the lower series in the graph symbolizes the coefficients extracted using the distance transforms method. As can be seen in Figure 2 from an anecdotal viewpoint, the privatized synthetic data series is an augmented outline of the noisy Fisher-Iris series used in the generation of the privatized synthetic data. The statistical analysis will further give more details that the statistical skeletal structure of the original data was maintained in the privatized synthetic data set. In Figure 3 the left sub-plot symbolizes the descriptive statistics of the original data, while the center sub-plot illustrates the statistical characteristics of noisy Fisher-Iris data, and the right sub-plot demonstrates the statistical traits of the generated privatized synthetic data. As shown in Figure 3 the statistical skeletal structural of the noisy data is maintained in the privatized synthetic data. For instance, the mean and median in the privatized synthetic data could be viewed as an augmentation of the same values in the noisy data, thus a statistical morphologic and skeletal structure of the original data. Since we derived the noisy data set by perturbing the original data, and likewise used distance transforms to extract coefficients from the noisy data and then generate the privatized synthetic data, the statistical skeletal structure of the original data is preserved, in this case, some level of augmentation has take place.

 ISBN: 978-1-61208-390-2

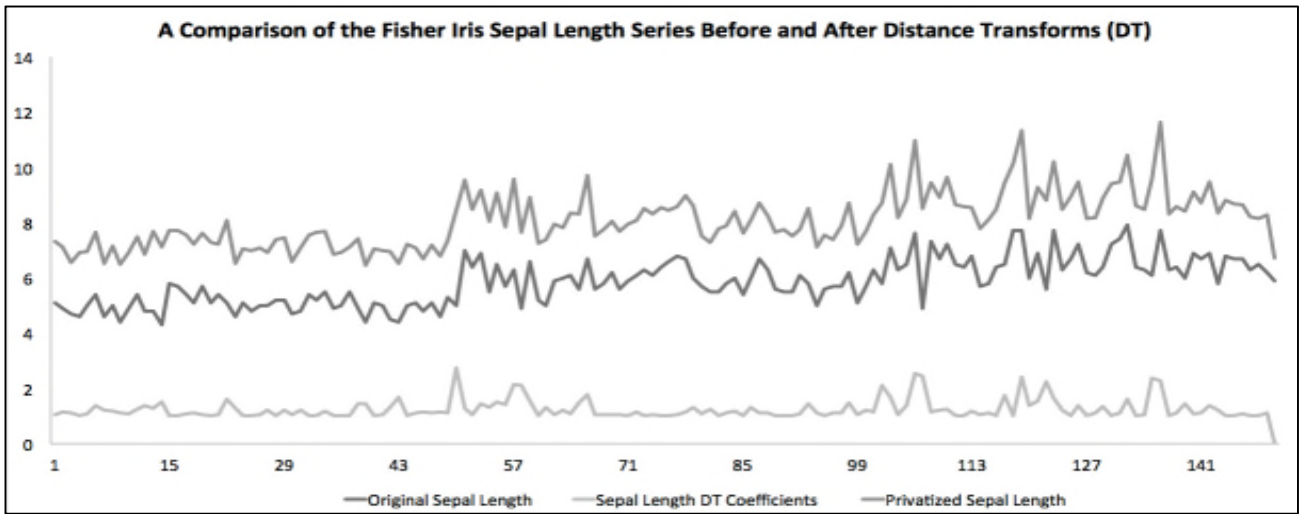


Figure 2: Original and Privatized Synthetic Fisher-Iris Sepal data series.

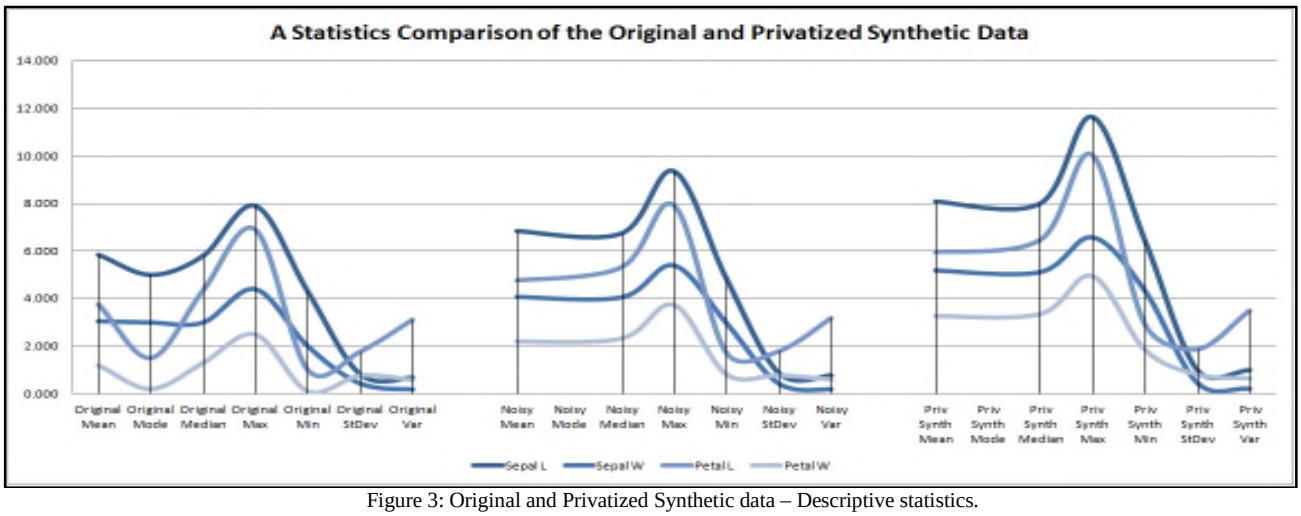


Figure 3: Original and Privatized Synthetic data – Descriptive statistics.

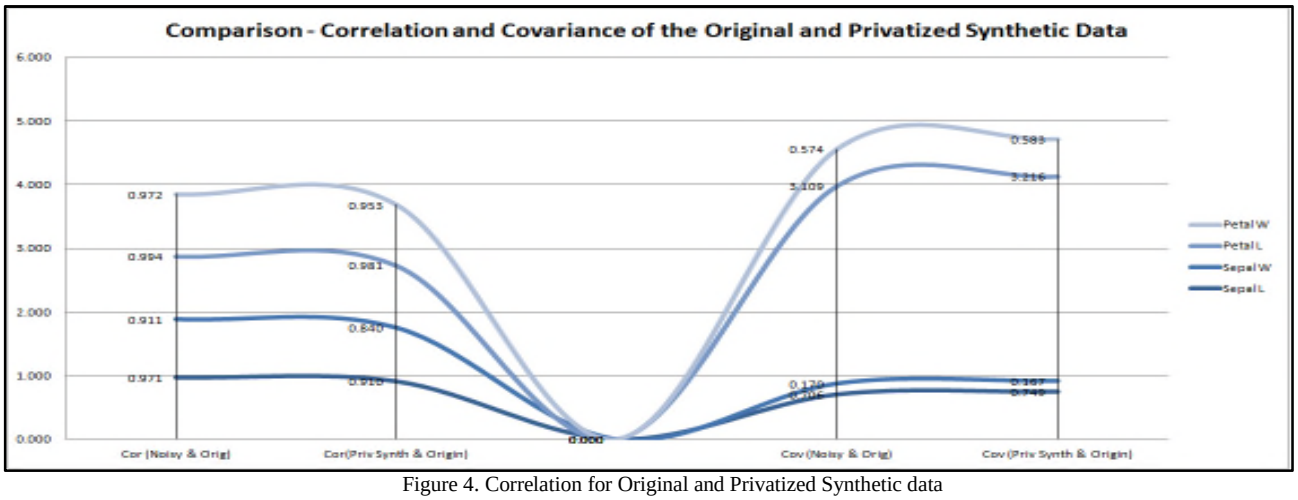


Figure 4. Correlation for Original and Privatized Synthetic data

 ISBN: 978-1-61208-390-2

For example, the mean values for the Sepal length is 5.834 in the original data, and 6.744 for the noisy data, with a dissimilarity of about 0.91. On the other hand, the mean value of 8.078 was registered for the privatized synthetic data, a dissimilarity of about 1.33 and 2.44, when compared with the noisy and original data respectively. The same statistical skeletal structure of the original, noisy, and privatized synthetic data sets is shown in Table 1 describing the descriptive statistics of the data sets. The same outcome is repeated for the other descriptive statistics, when comparing the original, noisy, and privatized synthetic data sets. It could then be argued that the privatized synthetic data could offer some level of data usability to researchers since it preserves some of the statistical characteristics of the original data – in this case, a statistical morphological and skeletal structure of the original data is preserved. The covariance results as shown in Figure 4 and Table 2 for the original, noisy, and privatized synthetic, vary between 0 and 3. For example, the Sepal length, Sepal width, and petal width, covariance varies between 0 and 1, showing a small proclivity for the data sets to grow together, despite the covariance being positive, in this case.

TABLE 1. ORIGIN AND PRIVATE SYNTHETIC DATA – DESCRIPTIVE STATISTICS

Statistics	Sepal L	Sepal W	Petal L	Petal W
Original Mean	5.843	3.054	3.759	1.199
Original Mode	5.000	3.000	1.500	0.200
Original Median	5.800	3.000	4.350	1.300
Original Max	7.900	4.400	6.900	2.500
Original Min	4.300	2.000	1.000	0.100
Original StDev	0.828	0.434	1.764	0.763
Original Variance	0.686	0.188	3.113	0.582
Noisy Mean	6.841	4.077	4.766	2.200
Noisy Median	6.744	4.060	5.323	2.333
Noisy Max	9.353	5.398	7.921	3.747
Noisy Min	4.846	2.978	1.716	0.819
Noisy StDev	0.883	0.433	1.784	0.779
Noisy Variance	0.780	0.188	3.183	0.607
Priv Synthetic Mean	8.078	5.185	5.959	3.279
Priv Synthetic Mode	#N/A	#N/A	#N/A	#N/A
Priv Synthetic Median	8.001	5.119	6.473	3.364
Priv Synthetic Max	11.631	6.576	10.028	4.962
Priv Synthetic Min	6.444	4.328	2.907	1.855
Priv Synthetic StDev	1.001	0.463	1.870	0.807
Priv Synthetic Var	1.001	0.214	3.497	0.651

However, for the Petal length, covariance registers a value of 3, indicating a possibility for the Petal length attributes in the data sets might grow together [9]. Additionally, as shown in Figure 4 and Table 2 the correlation values between the original, noisy, and privatized data sets vary from 0.840 to 0.994, approximately near +1, an indication of a better linear association and therefore a strong relationship [9]. For that reason, it could be argued that for the sake of data usability, the generated privatized synthetic data set might retain the statistical traits of the original data.

TABLE 2: CORRELATION FOR ORIGINAL AND PRIVATIZED SYNTHETIC DATA

Statistics	Sepal L	Sepal W	Petal L	Petal W
Cor (Noisy & Orig)	0.971	0.911	0.994	0.972
Cor (Priv Synth & Orig)	0.910	0.840	0.981	0.953
Cov (Noisy & Orig)	0.706	0.170	3.109	0.574
Cov (Priv Synth & Orig)	0.749	0.167	3.216	0.583

Clustering performance: Additionally, clustering analysis was done for the original and privatized synthetic data sets to further test for usability. Since the Euclidean distance was used in computing the morphological transforms of the privatized synthetic data set, Euclidean distance-based unsupervised learning methods, such as k-means clustering, could be used in testing for data usability of the privatized synthetic data sets.

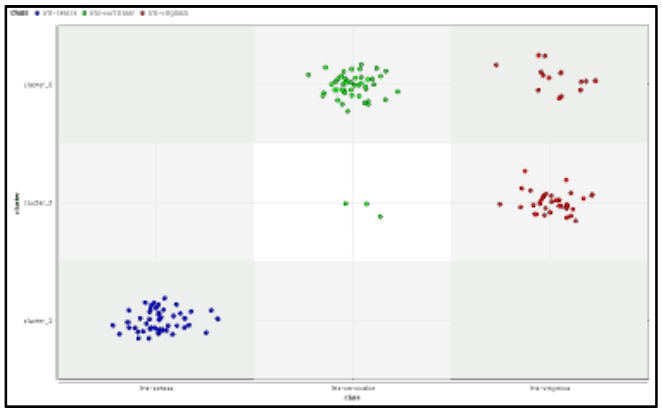


Figure 5. Original data clustering results

Furthermore, we put to test the suggestion by Mukherjee et al (2006), that one of the compensations of employing signal processing methods, such as discrete cosine, is that the Euclidean distance among points in the privatized data set could be preserved, with enhanced clustering outcomes [1][5][6]. In this case, we test to see if this proposal could hold when using image processing technique of distance transforms, as per our implementation. Clustering outcomes of the original Fisher-Iris data are shown in Figure 5 with the *x*-axis symbolizing the three classes – Iris Setosa, Iris Versicolar, and Iris Virginica; the *y*-axis symbolizes the number of clusters generated. K-means with Euclidean distance algorithm was employed for the clustering test, with k =3. An anecdotal view point of Figure 5 shows that Iris Virginica category did not cluster well for the original data. Yet still, from an anecdotal view, there seems to be an observable improvement in clustering results for the privatized synthetic data as shown in Figure 6, with the exception of the Iris Virginica attribute. However, after application of the Davis-Bouldin Index metric [10][11], to test the clustering performance, there was an actual

 ISBN: 978-1-61208-390-2

degradation in the clustering performance, as shown in Figure 8 and Table 3. The Davis Bouldin Index for the original data was reported at 0.668 and 0.765 for the privatized synthetic data. The lower the Davis Bouldin Index, the greater the clustering performance. Therefore the suggestion that signal processing techniques, such as, discrete cosine transforms, could improve Euclidean distance-based clustering results, did not hold for the image processing technique of Distance Transforms, in this experiment [1].

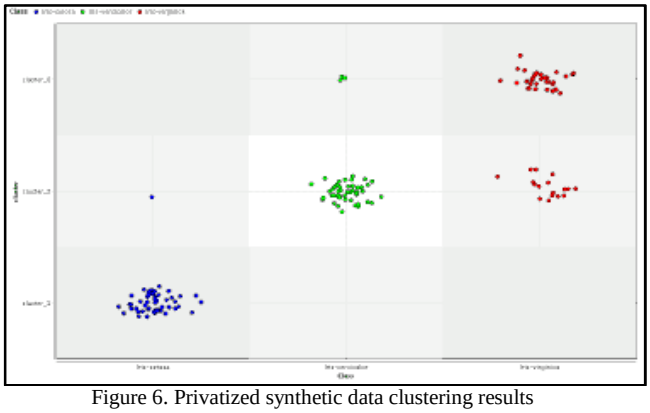


Figure 6. Privatized synthetic data clustering results

To mitigate this problem, we applied the Moving Average Filtering technique [11] on the privatized synthetic data set and then applied k-means clustering on the filtered data set again.

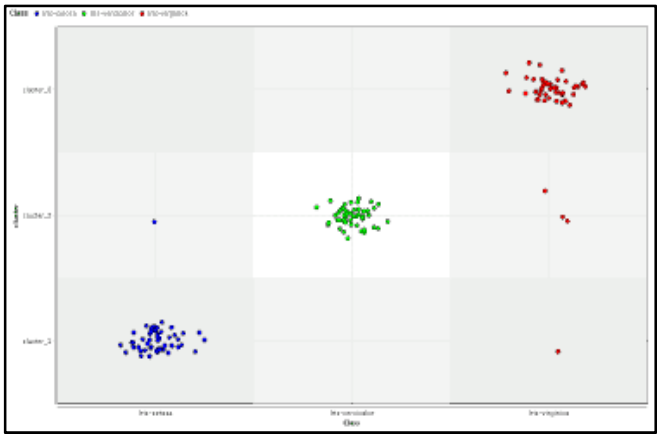


Figure 7. Filtered privatized synthetic data clustering results

Following the application of the moving average filter on the privatized synthetic data set, we clustered using k-means, with k=3, and as illustrated by the clustering outcome in Figure 7 there was an improvement with the Iris Virginica cluster. In fact, as illustrated in Figure 8, the Davis Bouldin index returned a value of 0.419 for the filtered privatized synthetic data, compared to the 0.668 for the original data, signifying an enhanced improvement in the clustering results for the privatized synthetic data, after filtering. Furthermore, using the distance between clusters metric – in this case, the average within centroid distance, to measure how well the clustering performed, Table 3 shows that the average within centroid distance of data points in the original data is approximately 0.5, while that for the non-filtered privatized synthetic data, is at 0.934. However, the average within centroid distance of data points in the filtered privatized synthetic data is about 0.477, an improvement that surpasses both the original and non-filtered privatized synthetic data sets.

We further tested for data usability by analyzing clustering performance, quantifying the number of items in each cluster, as a metric – the motivation was that the number of items in each cluster, in the privatized synthetic data should be similar to the number of items in each cluster, in the original data. For instance, as illustrated in Table 4 for the original data, there are 61, 50, 39, number of items in clusters 0, 1, and 2, in that order. However, for the non-filtered privatized synthetic data, there are 36, 49, 65, number of items in clusters 0, 1, and 2, respectively. Finally, for the filtered privatized synthetic data, we have the number of items as 46, 50, and 54, in clusters, 0, 1, and 2 respectively.

TABLE 3. CLUSTERING PERFORMANCE METRICS

Cluster Distance Performance	Original Data	Priv Synth Data	Filtered Priv Synth Data
Avg. within centroid distance	0.547	0.934	0.477
Avg. within centroid distance_cluster_0	0.562	0.657	0.635
Avg. within centroid distance_cluster_1	0.527	1.09	0.502
Avg. within centroid distance_cluster_2	0.492	0.961	0.268
Davies Bouldin Criterion	0.668	0.765	0.419

While the number of items in each of the clusters in the privatized synthetic data might not be close to that of the original data, we interpret this as a good indication of confidentiality, making it difficult for an adversary to know exactly how many items appeared in the clusters of the original data. Therefore, we could add to the argument that it might be possible to generate privatized synthetic data sets with acceptable levels of both confidentiality and usability.

TABLE 4. NUMBER OF ITEMS IN EACH CLUSTER

Cluster	Original Data	Synthetic Fisher Iris (DT) Data	Filtered Synthetic Fisher Iris (DT) Data
Cluster 0	61	36	46
Cluster 1	50	49	50
Cluster 2	39	65	54
Total	150	150	150

 ISBN: 978-1-61208-390-2

I. CONCLUSION

We have presented a confidential synthetic data generation heuristic, that employs a combination of data privacy and distance transforms techniques, for the generation of privatized synthetic data with similar statistical traits of the original data. We have also presented empirical results from applying unsupervised learning, using k-means, to test the usability of the privatized synthetic data sets. We applied average moving filtering on the privatized synthetic data and showed that filtering might help improve clustering results. Based on our empirical results from this study and implementation, we argue that it might be possible to generate privatized synthetic data sets, with acceptable levels of both privacy and data usability, while preserving the same statistical morphological and skeletal structure of the original, using a combination of data privacy, distance transforms, and filtering techniques. On the limitations of this study and future work, we focused on implementing the generation of privatized synthetic data sets using data privacy, distance transforms, and filtering techniques. While much effort could have been given to the testing of the privatized synthetic data sets to various adversarial attacks, our efforts were largely spent on the generation of the privatized synthetic data sets, leaving the study of attacks on privatized synthetic data sets for future work. Finally, generation of privatized synthetic data sets that retain the statistical structure of the original data, remains a challenge, and is in the early stages of research. More investigations on theoretical studies, practical implementations, and gathering of empirical results, is highly necessary for the advancement of privatized synthetic data set generation with enhanced levels of usability.

ACKNOWLEDGEMENT

Portion of this work was presented as part of the dissertation by the author in fulfillment of the requirements for the D.Sc. degree, in the Computer Science Department, at Bowie State University [10]. With great gratitude, I extend my sincere thanks Dr. Claude Turner and Dr. Soo-Yeon Ji, in the Computer Science Department, at Bowie State University, for the untiring assistance during this study. I would like to acknowledge with great appreciation, the HBGI Grant from the United States Department of Education for the facilitation of this work.

REFERENCES

[1] S. Mukherjee, Z. Chen, and A. Gangopadhyay, “A privacy-preserving technique for Euclidean distance-based mining algorithms using Fourier-related transforms,” VLDB J., vol. 15, no. 4, pp. 293–315, Aug. 2006.

[2] F. Y.-C. Shih and O. R. Mitchell, “A mathematical morphology approach to Euclidean distance transformation.,” IEEE Trans. Image Process., vol. 1, no. 2, pp. 197–204, Jan. 1992.

[3] O. Cuisenaire and B. Macq, “Fast Euclidean Distance Transformation by Propagation Using Multiple Neighborhoods,” Comput. Vis. Image Underst., vol. 76, no. 2, pp. 163–172, Nov. 1999.

[4] A. Rosenfeld and J. L. Pfaltz, “Distance functions on digital pictures,” Pattern Recognit., vol. 1, no. 1, pp. 33–61, 1968

[5] C. T. Huang and 0. Robert Mitchell, “A Euclidean distance transform using grayscale morphology decomposition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 16, no. 4, pp. 443–448, 1994.

[6] D. G. Bailey, “An efficient euclidean distance transform,” in LNCS Combinatorial Image Analysis, 2004, vol. 2, no. 3, pp. 394–408.

[7] D. G. Bailey, “Accelerating the distance transform,” in Proceedings of the 27th Conference on Image and Vision Computing New Zealand - IVCNZ ’12, 2012, pp. 162–167.

[8] K. Bache and M. Lichman, “Iris Fisher Dataset - UCI Machine Learning Repository.” University of California, School of Information and Computer Science., Irvine, CA, 2013.

[9] K. Mivule, “Utilizing Noise Addition for Data Privacy , an Overview,” in Proceedings of the International Conference on Information and Knowledge Engineering (IKE 2012), 2012, pp. 65–71.

[10] K. Mivule, "An Investigation of Data Privacy and Utility Using Machine Learning as a Gauge", D.Sc. Dissertation, Computer Science Dept., Bowie State University. 2014: 262 pages; ProQuest: 3619387.

[11] K. Mivule and C. Turner, "Applying Moving Average Filtering for Non-interactive Differential Privacy Settings", Procedia Computer Science, Volume 36, ISSN: 1877-0509, 2014, Pages 409-415. DOI: 10.1016/j.procs.2014.09.013

[12] S. Hajian and M. A. Azgomi, “A privacy preserving clustering technique using Haar wavelet transform and scaling data perturbation,” in 2008 International Conference on Innovations in Information Technology, 2008, pp. 218–222.

[13] O. Chertov and D. Tavrov, “Providing Group Anonymity Using Wavelet Transform,” in Lecture Notes in Computer Science - BNCOD 27, 2012, vol. 6121, pp. 25–36.

[14] N. V. Lalitha, G. Suresh, and P. Telagarapu, “Audio authentication using Arnold and Discrete Cosine Transform,” in 2012 International Conference on Computing, Electronics and Electrical Technologies (ICCEET), 2012, pp. 530–532.

[15] M. Niimi, F. Masutani, and H. Noda, “Protection of privacy in JPEG files using reversible information hiding,” in 2012 International Symposium on Intelligent Signal Processing and Communications Systems (ISPACS), 2012, no. Ispacs, pp. 441–446.

 ISBN: 978-1-61208-390-2

Flexible Manipulator Inspired by Octopus

Development of Soft Arms Using Sponge

Shunsuke Hagimori and Kazuyuki Ito
Dept. of Electrical and Electronics Engineering
Hosei University
Tokyo, Japan
e-mail: 10x2088@stu.hosei.ac.jp, ito@hosei.ac.jp

***Abstract*—In this study, we focus on the intelligent behavior of an octopus and describe the development of a flexible manipulator. By using the developed manipulator, we show that grasping behaviors similar to those of an octopus can be realized by the dynamics of the body without computation in its brain.**

Keywords-flexible manipulator; octopus; many degrees of freedom; grasping.

I. INTRODUCTION

In general, it is difficult to control a robot with many degrees of freedom. An advanced complex controller is required for controlling a robot. However, it is reported that some creatures—for example, a snake and an octopus—can exhibit complicated behavior such as locomoting, jumping, and grasping. A snake can locomote on rubbles using its many degrees of freedom [1]. Also, an octopus, for instance, is able to grasp various objects [2], [3], [4].

In our previous work to develop intelligent robots, we focused on the dynamics of the creature's body. In [5], we proposed a snake-like robot that has flexible joints. Due to the flexibility of the joints, the snake-like robot could work on rubbles without complex control.

In [6], we considered grasping by an octopus-like manipulator. An octopus has many degrees of freedom in its arms and can realize various intelligent behaviors. However, the brain of the octopus is very small, and how the octopus achieves intelligent behavior is currently an open question. Various studies about an octopus are currently being carried out [2], [3], [4]. Then, we focused on the dynamics of the arms of an octopus and hypothesized that some intelligent behaviors are realized by its dynamics instead of being controlled by its brain. We demonstrated that an octopus-like manipulator can grasp various objects without controlling each joint. However, as the links of the robot in [6] were rigid, it is difficult to grasp three-dimensional objects.

In this study, we extend our previous work. We develop a simple flexible manipulator for grasping three-dimensional objects on the basis of the fundamental behavior of the octopus [2] and demonstrate that the grasping task can be realized very easily without a controller.

This report consists of the following parts. Section II introduces the behavior of an octopus. Section III describes the proposed octopus-like manipulator. Section IV demonstrates that the proposed manipulator can grasp various unknown objects. Section V concludes the report.

II. FUNDAMENTAL BEHAVIOR OF AN OCTOPUS

It is reported that an octopus can stretch its arm from the root to the tip in sequence while keeping its curved shape, as shown in Figure 1(a). When the curved point contacts an object, the arm wraps around the object [2], [3], [4].

This fundamental behavior is effective in grasping unknown objects. In other words, if the arm stretches from the tip and the tip contacts the object first, the arm cannot wrap around the object, as shown in Figure 1(b).

The important question in this research is how we design the robot in order to realize octopus-like behavior. We focus on the dynamics of the flexible arm and hypothesize that the arm moves and wraps around the obstacle passively by simply adjusting the compliance of the muscles of the arm. In other words, the arm can grasp the unknown obstacle without a complex control signal from the brain [2], [3], [4], [6]. The grasping behavior is realized by the dynamics of the flexible arm [6].

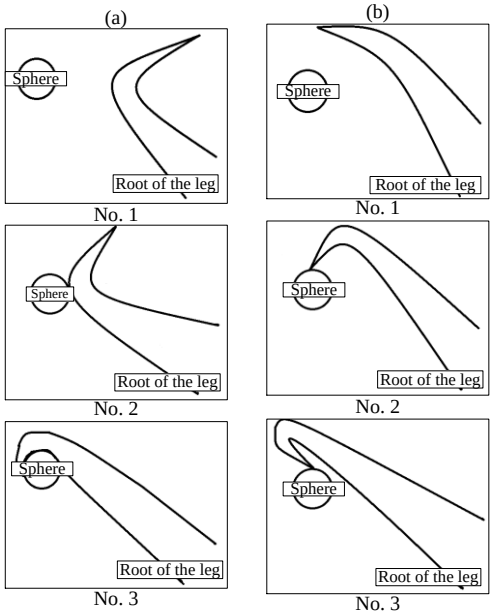


Figure 1. (a) Effective strategy of an octopus and (b) failure mode.

 ISBN: 978-1-61208-390-2

In this study, we developed an octopus-like manipulator with a simple mechanism and demonstrated that this fundamental behavior can be realized by the dynamics of a flexible material and the constraints of wires.

III. OCTOPUS-LIKE MANIPULATOR

To realize the fundamental octopus behavior, we developed a simple flexible manipulator. Figures 2–7 show the mechanism of the manipulator. The manipulator consists of the sponge arms, the rubbers that simulate the compliance of the muscles, and wires to move the arms. A motor can pull the wire via pulleys, and the wires close the arms.

Note that the rubber is bonded to the sponge arm as it is stretched, as shown in Figure 2. In addition, the length of the stretching is increased from the root to the tip. From the tension of the rubber, the arm is twisted to a natural position, as shown in Figure 3.

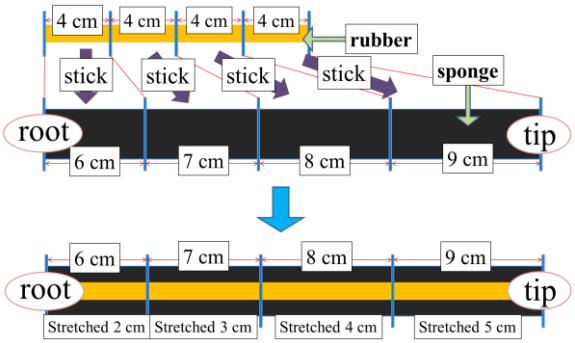


Figure 2. Rubber and sponge for an arm.

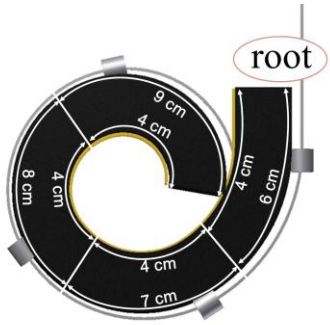


Figure 3. Natural position (open position) of an arm.

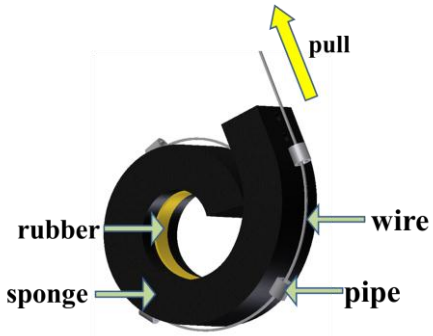


Figure 4. Structure of an arm.

With the asymmetric tension (compliance), the shape of the arm is also asymmetric. The wires are pulled by rotating a motor, and the arms move from the root to the tip like an octopus. Because the arms are flexible, the shapes of the arms adapt to the unknown objects. Figure 5 shows the developed manipulator. We designed the manipulator to grasp an object from the upper part. Active pulleys with a motor are placed in the upper aluminum frame. Figure 6 shows the top view of the arms. Three arms are installed under the pulleys via an aluminum frame at a 120° interval. Figure 7 shows the mechanism to move the arms. The wires are pulled by rotating the motor, and the three arms are closed at the same time. Then, the manipulator can grasp an object.

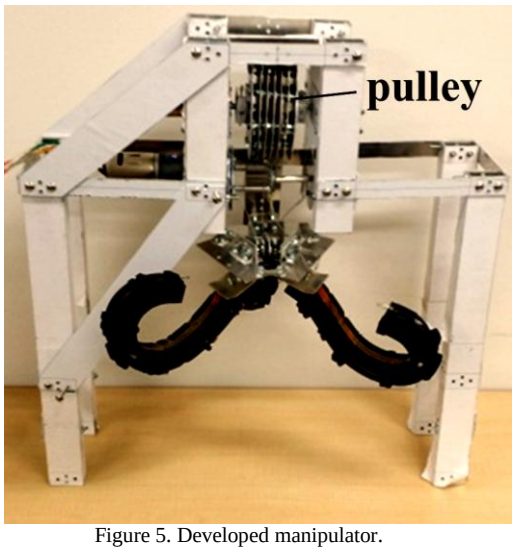


Figure 5. Developed manipulator.

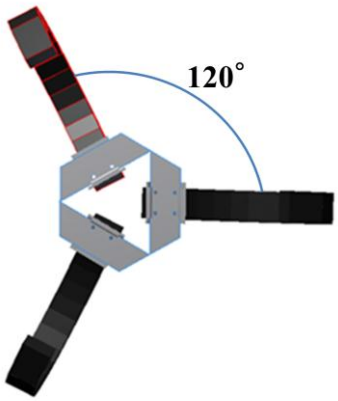


Figure 6. Top view of the arms.

 ISBN: 978-1-61208-390-2

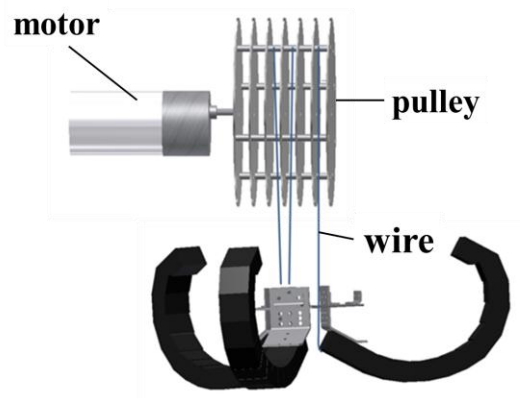


Figure 7. Mechanism to move the arms.

IV. EXPERIMENT

Figures 8 and 9 show examples of the results.

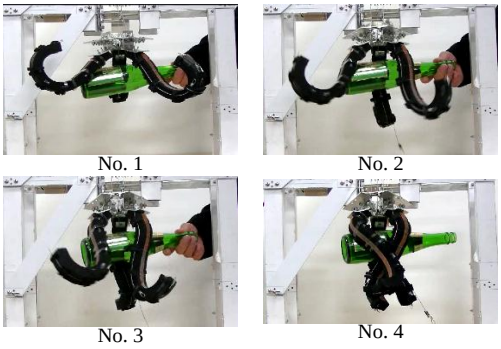


Figure 8. Structure of the manipulator.

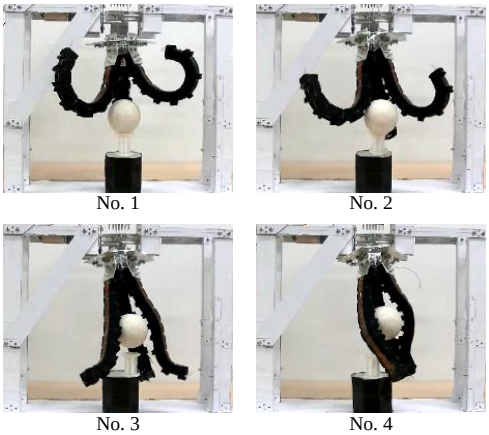


Figure 9. Experimental results of grasping a ball on the stand.

In this experiment, the manipulator worked by simple operation. An operator just turns ON the switch of the motor, and the manipulator can grasp an object. Then, the operator turns OFF the switch.

In this way, the manipulator could work only when turning the switch of the motor ON/OFF.

We confirm that the developed manipulator realizes octopus-like motion and can grasp various unknown objects. In these experiments, we simply pulled the wires using the motor, and no complex control signals were required. The grasping behavior was realized passively by the dynamics of the body.

In general, it is reported that much computation cost is required to control a robot with many degrees of freedom [7]. In contrast, the results of this research show that the proposed manipulator requires no computational cost. Necessary computation for controlling the arms was conducted by the dynamics of the arms instead of a computer.

V. CONCLUSION

In this study, we focused on the intelligent behavior of an octopus. We developed a flexible manipulator that realized octopus-like grasping behavior, and we showed that the grasping of unknown objects could be realized by the dynamics of the body without complex control signals.

ACKNOWLEDGMENT

This research was partially supported by the Japan Society for the promotion of science through the Grant-in-Aid for scientific research (C) 24500181.

REFERENCES

[1] J. Gray, “The Mechanism of Locomotion in Snakes,” Exp. Biol., Vol. 23, pp. 101-123, 1946.

[2] G. Sumbre, Y. Gutfreund, G. Fiorito, T. Flash, and B. Hochner, “Control of Octopus Arm Extension by a Peripheral Motor Program,” Science, vol. 293, no. 5536, pp. 1845-1848, 2001.

[3] Y. Yekutieli, G. Sumbre, T. Flash, and B. Hochner, “How to Move with No Rigid Skeleton?” Biologist, vol. 49, no. 6, pp. 250-4, Dec. 2002.

[4] Y. Gutfreund, T. Flash, G. Fiorito, and B. Hochner, “Patterns of Arm Muscle Activation Involved in Octopus Reaching Movements,” J. Neurosci., vol. 18, no. 15, pp. 5976-5987, Aug. 1998.

[5] K. Ito and R. Murai, “Snake-Like Robot for Rescue Operations—Proposal of a Simple Adaptive Mechanism Designed for Ease of Use,” Advanced Robotics, vol. 22, no. 6-7, pp. 771-785, 2008.

[6] S. Kuroe and K. Ito, “Autonomous Control of Octopus-Like Manipulator Using Reinforcement Learning,” In Sigeru Omatu, Juan F. DePaz Santana, Sara Rodríguez-González, José M. Molina, Ana M. Bernardos,Juan M. Corchado Rodríguez, editors,Distributed Computing and Artificial Intelligence Advances in Intelligent and Soft Computing, vol. 151, pp. 553-556, 2012.

[7] Y. Zhang and R. P. Paul, “Robot Manipulator Control and Computational Cost,” *Scholarly Commons*. [Online]. Available from: http://repository.upenn.edu/cis_reports/621 Feb. 1988.

 ISBN: 978-1-61208-390-2

A Method to Understand Psychological Factors Needed to Improve Learning Behavior

Yuto Omae, Katsuko T. Nakahira, and Hirotaka Takahashi
Nagaoka University of Technology
Niigata, Japan
e-mail: y_omae@stn.nagaokaut.ac.jp, katsuko@vos.nagaokaut.ac.jp, hirotaka@kjs.nagaokaut.ac.jp

***Abstract*— Learning behavior is influenced by psychological factors. Therefore, if teachers desire to improve the learning behavior of their students, they need to know the relevant psychological factors and their role in improving learning behavior. From this point of view, this paper reports a method to quantitatively understand the psychological factors needed to improve learning behavior and their shortages by using a decision tree. Our proposed method is expected to effectively utilize adaptive learning for class design to improve students' learning behavior.**

Keywords-Psychometrics; Data mining; Learning behavior.

I. INTRODUCTION

We propose a method to quantitatively understand psychological factors related to improving learning behavior.

In cognitive psychology, knowledge is categorized as procedural or declarative [1]. "Procedural knowledge" is knowledge about performing various actions (e.g., a calculating ability such as addition, subtraction, division or multiplication). It is acquired by repetition of its action. "Declarative knowledge" has a network structure that regards knowledge as a "node" and the relations of knowledge as "edges". One node connects to another along an edge [1]. To acquire declarative knowledge, the learner needs to learn while thinking about the meaning of each bit of knowledge. From the above, if the teacher wishes students to acquire procedural knowledge, he/she has to assign a task or homework involving exercise with repetition (e.g., in the case of calculating ability, the teacher assigns many numerical calculations to students). In contrast, for students to acquire declarative knowledge, the teacher must both assign an appropriate task and appropriate learning while thinking about the meaning.

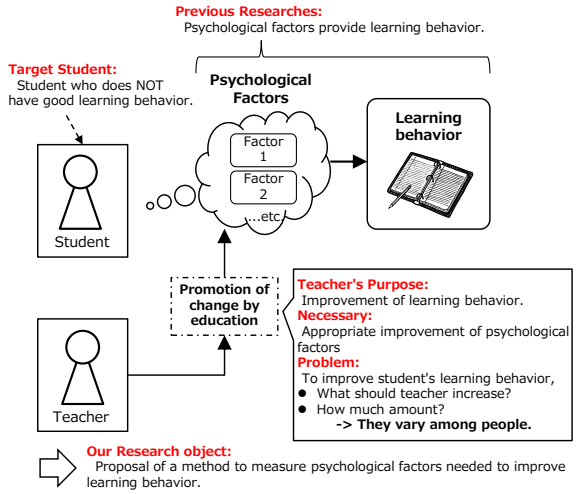


Figure 1. Problem to occur when teacher improves student's learning behavior.

However, Teranishi pointed out that the number of people who learn without thinking about it is increasing [2]. If the teacher wishes for them to acquire the appropriate declarative knowledge, he/she has to improve their learning behavior. However, changing learning behavior is difficult [3]. Thus, we consider that teaching students declarative knowledge is more difficult than teaching them procedural knowledge. For the above reason, our research target is to improve learning behavior needed to acquire declarative knowledge.

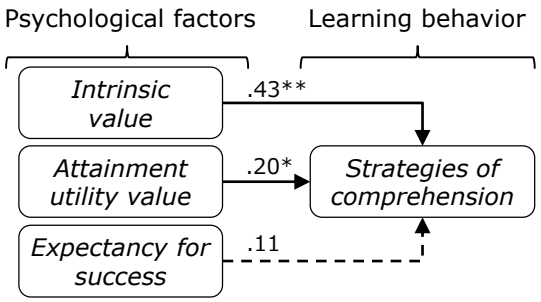
Horino mentioned the importance of improving the psychological factors that provide learning behavior [3]. Previous research [3][4][5] also mentioned that the effective factors related to learning behavior are psychological factors. Therefore, to improve a student's learning behavior, it is necessary for the teacher to improve the student's psychological factors. Figure 1 illustrates a problem that occurs when a teacher improves a student's learning behavior. The teacher is going to improve the student's learning behavior by his/her education. According to previous research on improving learning behavior, it is necessary to improve the psychological factors. However, the required psychological factor to improve learning behavior and its shortage vary among people. Teachers need to understand the missing psychological factors and their shortages. However, there is presently no method to measure a psychological factor to improve students' learning behavior. Because of this, the following problem occurs. The teacher cannot understand what he/she should change in the student's psychological factors and by what amount it should increase to improve the student's learning behavior. To solve this problem, this paper proposes a method to quantitatively understand the psychological factors required to improve a student's learning behavior.

In Section II, we present the necessary strategies to achieve our purpose and an outline for it. In Section III, we detail the results of a survey about psychological factors affecting learning behavior. In Section IV, we explain a

 ISBN: 978-1-61208-390-2

TABLE I. TECHNICAL TERM

Psychological factors
Intrinsic value means fun and interest about learning contents.
Attainment utility value means recognitions of merit about acquiring and using knowledge of learning contents.
Expectancy for success means confidence about learning contents.
Learning behavior
Strategies of comprehension are learning behavior such as study with understanding meaning.



	Mean	SD
Intrinsic value	0.50	0.29
Attainment utility value	0.64	0.22
Expectancy for success	0.57	0.21
Strategies of comprehension	0.57	0.20

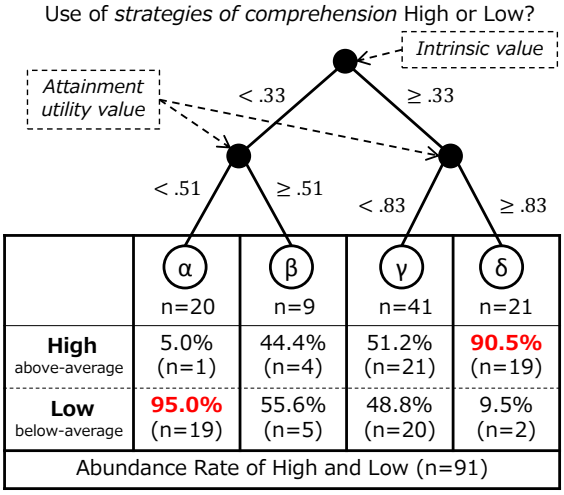
Figure 2. SEM of psychological factors and learning behavior. Dashed line is NOT statistically significant.

method to quantitatively measure the psychological factors needed to improve learning behavior. In Section V, we provide this paper's summary and the prospects for future work.

II. OUTLINE

We focus on mathematics as the subject of our method because Ichikawa says that many learners are weak in this area [6]. We adopted *strategies of comprehension*, which are effective learning strategies for knowledge acquisition in mathematics (defined in Learning Behavior in Table I) [2]. Ichihara tried to explain that the use of *strategies of comprehension* is influenced by the elements in Eccles's expectancy-value theories (*Intrinsic value*, *attainment utility value*, and *expectancy for success*) [4][7]. In imitation of this, we adopted *intrinsic value*, *attainment utility value*, and *expectancy for success* as our psychological factors (Psychological Factors are defined in Table I). Based on the above discussion, this paper focuses on learners who use few *strategies of comprehension*. Here, we describe our method to understand quantitatively the psychological factors needed to increase the use of *strategies of comprehension*. In order to do this, we need to address the following two points.

(1) Understanding the kind of psychological factor that provides use of *strategies of comprehension*.



Dependent variable (Binary variable, "High" or "Low") :
→ *"Strategies of comprehension"*
Independent variable (Quantitative variable, [0,1]) :
→ *"Intrinsic value", "Attainment utility value"*

Figure 3. Decision tree.

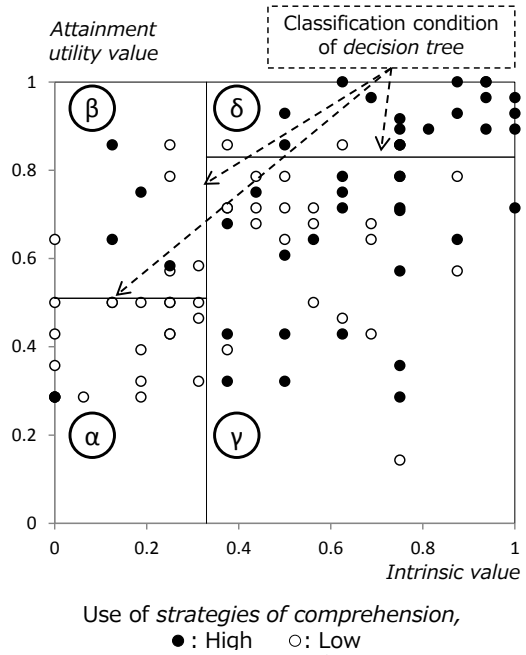


Figure 4. Scatter plot to be classified by decision tree.

(2) Understanding the conditions (in the form of the *Decision Tree* discussed in Section III) of the psychological factors selected in (1) to encourage use of *strategies of comprehension*.

Item (1) is covered in previous research [4]. Thus, we re-inspect it using *Structural Equation Modeling* (*SEM*). We apply item (2) to the result of clustering by *decision tree* based on the result of (1). We then consider conditions connected to improving the use of *strategies of comprehension*. To get the full picture of items (1) and (2), we surveyed 91 students of high school to score items on their *intrinsic value*, *attainment utility value*, *expectancy for*

TABLE II. AN EXAMPLE OF USED QUESTIONNAIRE. THERE ARE ALL ITEMS IN [4].

***Intrinsic value* (Total 7 items)**
(1) I think mathematics is interesting. (2) I like mathematics. (3) I enjoy studying mathematics. (In addition, there are 4 items.)
***Attainment utility value* (Total 8 items)**
(1) It is important for me to be good at mathematics. (2) I think the knowledge of mathematics will be useful in future. (3) The knowledge of mathematics is important for learning other subjects. (In addition, there are 5 items.)
***Expectancy for success* (Total 7 items)**
(1) I have confidence about being good at mathematics. (2) I have confidence about understanding learning contents in mathematics lesson. (3) I have confidence about getting good score in the mathematics tests. (In addition, there are 4 items.)
***Strategies of comprehension* (Total 7 items)**
(1) When I study mathematics, I prove a theorem. (2) When I read mathematics' textbook, I use critical thinking (e.g., Why is this theorem proved by this proof process?). (3) When I solve mathematics' problem, I think specific image. (In addition, there are 4 items.)
Answer Form (6-Likert-Scale)
About the above question items, please choose one : 1: Very **Negative** 2: **Negative** 3: Little **Negative** 4: Little **Positive** 5: **Positive** 6: Very **Positive**

success, and *strategies of comprehension*. We created the survey questions based on Ichihara's items, generated by a factor analysis (Table II shows some of them. The answer form is on a 6-Likert scale).

III. RESULTS AND DISCUSSIONS

We standardized *intrinsic value*, *attainment utility value*, *expectancy for success*, and *strategies of comprehension* to range from 0 to 1. Figure 2 presents the result of (1), along with the mean and standard deviation. The partial regression coefficients of *intrinsic value* and *attainment utility value* were statistically significant. This result is similar to previous research [4]. We thus determined that *intrinsic value* and *attainment utility value* could be used to categorize the amount of *strategies of comprehension* used. For the above reason, we performed clustering in the form of a *decision tree*. To do this, *strategies of comprehension* was classified into above average (High) and below average (Low). We then regarded *strategies of comprehension* as the dependent variable and *intrinsic value* and *attainment utility value* as independent variables. Based on these conditions, we performed clustering in the form of a *decision tree* (see Figure 3 and the corresponding scatter plot in Figure 4). We set the horizontal axis as the *intrinsic value*, the vertical axis as the *attainment utility value*, and the density of colors as the abundance ratio of *strategies of comprehension* state in terms of "High (Black)" and "Low (White)", and

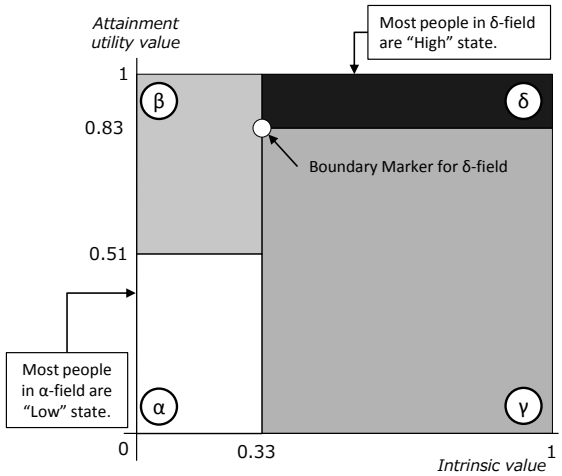


Figure 5. 2D Map of Psychological Factors and Learning Behavior. The density of color means High and Low rate. Black=High, White=Low.

constructed a *2D Map of Psychological Factors and Learning Behavior* (Figure 5). The map has four fields according to the abundance ratio ("High" or "Low") of the amount of *strategies of comprehension* used and the value of the psychological factors.

- **α-field (White)** Most of the subjects are in a "Low" *strategies of comprehension* state. *Intrinsic value* and *attainment utility value* are low.
- **β-field (Grey)** The abundance ratio of the *strategies of comprehension* state: ("High" or "Low") is nearly half-and-half. *Intrinsic value* is low.
- **γ-field (Grey)** The abundance ratio of the *strategies of comprehension* state ("High" or "Low") is nearly half-and-half. *Attainment utility value* is low.
- **δ-field (Black)** Most of the subjects are in a "High" *strategies of comprehension* state. *Intrinsic value* and *attainment utility value* are high.

If a student's *strategies of comprehension* state is "Low," it is desirable to move the point of their *intrinsic value* and *attainment utility value* into the δ-field to improve their use of *strategies of comprehension*.

IV. PROPOSED METHOD

Based on the result in Section III, we determined that the shortages of *intrinsic value* and of *attainment utility value* to improve the use of *strategies of comprehension* represent the difference between "Student Marker" and "Boundary Marker for δ-field" (Figure 6). The teacher can understand the shortage of these psychological factors, which vary from person to person, to improve learning behavior by following five steps.

Step 1. Measure *strategies of comprehension* of students using the questionnaire in Table II.

 ISBN: 978-1-61208-390-2

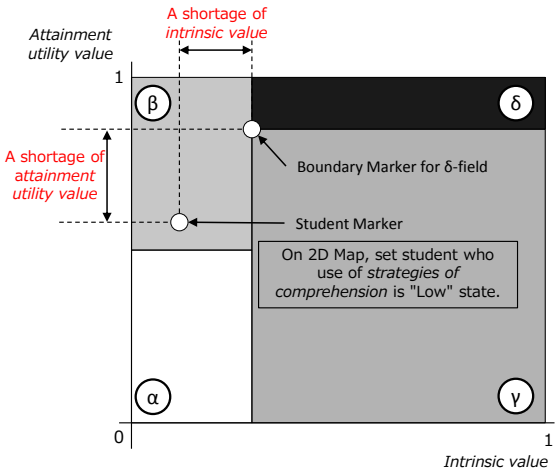


Figure 6. Definition of the degree of psychological factor needed to improve use of strategies of comprehension.

Step 2. Sample students in the "Low" state for *strategies of comprehension*.

Step 3. Measure *intrinsic value* and *attainment utility value* of these students (using the questions in Table II).

Step 4. Understand the shortages of *intrinsic value* and/or *attainment utility value* by calculating the difference between Student and Boundary Marker for the δ-Field on a 2D Map (Figure 6).

Step 5. If the teacher wants to know the average shortages for his/her own class, he/she calculates them from the individual shortages.

We applied our proposed methods to two classes (*a* and *b*). Table III presents the results of Steps 1 to 5. The insufficient psychological factors of *a-Class* were *intrinsic value* and *attainment utility value*. However, the only insufficient psychological factor of *b-Class* was *attainment utility value*. With this method, we can quantitatively understand the psychological factors needed to improve the use of *strategies of comprehension*.

V. SUMMARY AND FUTURE WORK

There presently is no method to measure the psychological factors needed to improve learning behavior. Thus, this paper proposed a measuring method for these factors. The adopted psychological factors were *intrinsic value* and *attainment utility value*. The adopted learning behavior was *strategies of comprehension*. We proposed a method to understand the shortage of each psychological factor in order to improve learning behavior using these factors. This paper described the research process as follows.

(1) Re-inspect the causal relationship reported in previous research (Figure 2) [4].

(2) Construct a *2D Map of Psychological Factors and Learning Behavior* using a *decision tree* (Figure 5).

(3) Determine the shortages of the psychological factors needed to improve the use of *strategies of comprehension* (Figure 6). (The procedures to achieve this are given as Steps 1 to 5 in Section IV).

TABLE III. RESULT OF STEPS 1-5

	Average shortage amount	
	Intrinsic value	*Attainment utility value*
a–Class (*n*=22)	0.16	0.36
b–Class (*n*=25)	0.01	0.35

From (2) and (3), a teacher can quantitatively understand the shortages of psychological factors that vary among people to improve students' learning behavior. We applied this method in two classes (Table III). In the results, the shortages of the psychological factors were different in each class (Table III). By using this result, teachers will be able to design adaptive learning approaches based on improving learning behavior using psychological factors.

As future work, we will perform education to increase *intrinsic value* and *attainment utility value* in two classes (*a* and *b*, Table III) while monitoring their psychological factors on a *2D Map of Psychological Factor and Learning Behavior* time-serially. Based on this, our next step will be to improve their use of *strategies of comprehension*.

ACKNOWLEDGMENT

This work was supported in part by JSPS KAKENHI Grant-in-Aid for Scientific Research (No. 24501146). This work was also supported in part by Nagaoka University of Technology Presidential Research Grant (D). We would like to thank Takako Mitsui, Yoko Tsuchiya and Rai Shukuin at Yamanashi Eiwa Junior and Senior High School for cooperating on our research.

REFERENCES

[1] E. D. Gagne, C. W. Yekovich, and F. R. Yekovich, "The Cognitive Psychology of School Learning", Little Brown. 1993.

[2] Y. Teranishi, "How High-school Students Think of Formulas and Theorem in Mathematics -Correlation with Learning Beliefs, Strategies, and Performance-" The bulletin of the Graduate School of Education of Waseda University. 16 (1), 2008, pp. 1-13.

[3] M. Horino, and S. Ichikawa, "Learning Motives and Strategies in High-School Student's English Learning" The Japanese Journal of Educational Psychology. 45 (2), 1997, pp. 140-147.

[4] M. Ichihara, and K. Arai, "Moderator Effects of Meta-Cognition: A Test in Math of a Motivational Model" The Japanese Journal of Educational Psychology. 54 (2), 2006, pp. 199-210.

[5] P. R. Pintrich, and E. V. DeGroot, "Motivational and Self-regulated Learning Components of Classroom Academic Performance" Journal of Educational Psychology, 82, 1990, pp. 33-40.

[6] S. Ichikawa, "Ninchi kaunseringu kara mita gakusyu houhou no soudan to shidou (Counseling and Guidance of the Learning Method from Cognition)" Brain Publisher, 1998.

[7] J. E. Parsons, T. F. Adler, R. Futterman, S. B. Gof, C. M. Kaczala, J. L. Meece, and C. Midgley, "Expectancies Values and Academic Behaviors. In J.T.Spemce(Ed)" Achievement and Achievement Motivation. San Francisco, CA: Freeman, 1983, pp. 75-146.

 ISBN: 978-1-61208-390-2

Ecology of Spam Server Under Resilience Force in the e-Network Framework

Katsuko T. Nakahira, Kakeru Yamaguchi, and Muneo Kitajima

Nagaoka University of Technology
Niigata, Japan
Email:katsuko@vos.nagaokaut.ac.jp,yukidaruma1232@yahoo.co.jp,mkitajima@kjs.nagaokaut.ac.jp

***Abstract*—We propose the concept "*ecology of a spam server*," based on the *e*-Network and *resilience* in an effort to clarify its mechanism, which will contribute to the development of security and network strategies. We consider the microstructure of *resilience* with the *e*-Network framework and demonstrate three stages of spam servers: secure but underlying, developing, and critical. Using these features, we introduce the Evolution Diagram (ED), a method for quantitatively representing the patterns of the evolution history of a server. From this diagram, we derive three indexes to measure resilience: maximum transmission potential, continuity, and reproducibility. Through these indexes, we define resilience, which is divided into eight classes depending on the existence/absence of these three primitive features. We calculate the *resilience* of the individual spam servers. This idea would lead to useful tools for producing strategies not only for security but also for Internet/country domain governance.**

Keywords–ecology of spam server, resilience, e-Network, ED transition diagram

I. INTRODUCTION

In this study, we develop the concept "ecology of a spam server" in an effort to understand the mechanism under which spam evolves and spreads. In addition, we determine how it is related to the *e-Network* [1].

Currently, spam is considered a principal factor that may cause serious problems (e.g., decreasing one's productivity due to the slip of important mail and increasing the cost of administrating mail substratum). Many studies have been conducted and a number of white papers have been published concerning security-related spam. These studies have proposed a variety of anti-spam strategies. For example, Graham [2] addressed the following issues based on the userprovider configuration and networking scheme with some descriptions of advantages, demerits, and roles of various categories: filtering (signature-based/Bayesian (statistical) rule-based (heuristic)/challenge-response), secret address, junk address, penny per mail, mail server blacklists, filters that fight back (FFBs), slow senders, laws, and complaints to spammers' Internet Service Provider (ISP)'s. Recently, other strategies have been developed. Li and Hsieh [3] developed a group-based anti-spam framework. They analyzed community behavior of spammers through a large collection of spam mail to identify structures of spammers using spam traffic data collected on a domain mail server. They suggested that the number of members in a group and the number of groups with which a spammer is associated are useful measures for developing group-based anti-spam strategies. Stanković and Simič [4] proposed effective strategies for defending against botnets, consisting of a list of measures and activities along with some explanatory descriptions. Van Staden and Venter [5] proposed anti-spam strategies for detecting botnet activities and tracing botmasters.

Expanding this research, the objective of the present study is to construct a method for estimating Internet governance for each region in the world. To do this, we assume that the spam server exists as social ecology. Thus, we must understand spam servers' behavior in the context of *human* involvement. In this paper, we introduce two concepts, the *e*-Network and *resilience*, to clarify the *ecology of* a spam server. By definition, spam servers distribute spam mail. However, the spam server is not installed as a "spam server" from the outset. It "behaves" as a spam server at the moment but may return to acting as a "normal" server at any time in the future. However, it may continue to behave as a spam server, or even become a worse spam server. How the behavior of a spam server changes over time depends on several technological, human, and social factors. This study regards this phenomenon as the ecology of a spam server. This paper discusses dynamics in the *e*-Network framework [1] and proposes *resilience* as a primary force that shapes the behavior of spam servers.

e-Network. This study assumes that the ecology of a spam server should emerge from interactions among the factors defined in the *e*-Network framework proposed by [1]. Table I lists the fundamental components in the *e*-Network: human factor, substratum factor, products factor, and environment factor. The media, which are restricted to spam mail in this study, connect them. The following are examples of interaction between factors:

- A user who intends to send spam mail must use PCs and ISPs (human-substratum interaction).
- A user sends spam mail to earn money (human-environment interaction).
- The registry must control Top Level Domain (TLD)s (substratum-environment interaction).

Resilience force. *Resilience*is defined as "the ability of a network administer and maintain an acceptable level of service in the presence of various faults and [such] challenges to normal operations" in the context of network risk management, focusing on the relationship between security and *resilience* (e.g., see [6] and [7]). This study, however, reinterprets it within the *e*-Network framework: a server that has become a spam server might revert to a normal secure server as a result of interactions between the substratum factor (server and network) and the

 ISBN: 978-1-61208-390-2

TABLE I. THE FRAMEWORK COMPONENTS IN E-NETWORK

component	role	variable
human factor	*human behavior*	user
substratum factor	*the device for execution*	client, server, the Internet
environment factor	*surround human and substratum*	law, freedom of speech, education, income, a custom, etc.
product	*produced from interaction with human and substratum factor*	information contents
media	*connection device between the relation of all components*	language, images, etc.

human factor, environment factor, and/or products factor. Some interactions could be strong enough to make an insecure server revert to a secure one, whereas others could be too weak to make this happen. This paper metaphorically considers that these interactions work as a source of forces, called *resilience* force here.

Resilience force characterizes the state of a spam server over time. It could be secure, developing, or critical. If the *resilience* force of a server is strong, it allows the server to revert to its secure state even if the server begins sending spam mail. If the *resilience* force of a server is weak, a server that has begun to send spam mail is likely to develop the spam-sending activity to a critical state.

In the following sections, this paper regards the phenomenon that a spam server changes its state over time as the evolution of a spam server, and proposes the *Evolution Diagram (ED)* for quantitatively representing the pattern of a spam server's evolution. An ED value, which is a cumulative and integral value characterizing the evolution history of a server, and its derivatives (e.g., reproducibility, continuity, and potential) are candidates for expressing how *resilience* force has worked on the server. In section 2, we introduce three ecological stage of a server in the *e*-Network. In section 3, we develop the method how to measure strengths of *resilience* force with ED. In section 4, we construct ED transition diagram and relate to *resilience*. In section 5, we discuss two ecological scenarios for spam server. In section 6, we show the results of observed ecology for spam servers.

II. THREE ECOLOGICAL STAGES OF A SERVER IN THE *e*-NETWORK

A server can be in one of the three ecological stages in the e-Network: secure but underlying, developing, or critical. Each of these stages is closely related to how the *resilience* force works in the situation defined by the detailed conditions of the *e*-Network around the server. Using Fig. 1, this section explains the conditions of the elements of the *e*-Network (human (users), substratum, and environment (government)) and their interactions, and introduces *resilience* force.

Stage 1: Secure but underlying. When a mail server installs or starts mail service, it should be secure. However, it could become a spam server just because it works as a mail server. Therefore, the status of a mail server at this stage is "secure and underlying" for several reasons. It is likely that a new mail server is equipped with the latest versions of OS and protocol. In addition, when a mail service is started, most users who want to use this service are moral users. The scale of service is just right for skilled administrators to run and maintain it appropriately. The range of mail usage is limited: users use the service simply to contact their friends, family, or business partners. They tend to reject its immoral use. For these reasons, Internet governance is maintained by the behavior of moral users.

Moving to the underlying state. As time goes by, the conditions that allow the server to keep its status secure may change, and some problems may arise. The first spam is generated. However, this process occurs only occasionally. For example, immoral users who join the service may send illegal spam e-mail (e.g., junk mail). In order to deal with the increased number of users, the service provider must reinforce the server's substratum (e.g., deploy new servers or increase transmission speed). As the number of servers grows and the transmission rate increases, the administrator's task of keeping the service secure becomes more complicated. Due to the shift of service users' and providers' structures, some mail servers may allow a large amount of junk e-mail to be sent, mainly because the administrator cannot distinguish good e-mail from junk e-mail.

Staying in the secure and underlying state. With strong *resilience* force, the service provider can stay in the secure but underlying state by reducing the possibility of immoral use through broadcast constraints, enhanced maintenance, and skilled administrators.

Moving to the developing state. With weak *resilience* force, a server moves to the developing stage, where it sends spam mail chronically. This stage is triggered by worsening behavior of users and/or the substratum. Some users may become involved in immoral behavior, such as developing immoral technology (e.g., virus software), hacking/cracking algorithms, and automatically sending spam programs. The server provides little maintenance, due to the lack of skilled administrators and a poor engineering staff.

Stage 2: Developing. A strong *resilience* force originating from the social system might produce several strategies to improve the service server conditions and cause it to revert to the secure and underlying stage. Otherwise, the number of immoral service users will increase, and immoral technologies will proliferate continuously and widely.

A strong *resilience* force originating from Internet governance results in increased control over these immoral behaviors through laws or strict governance. These interventions involve legal force to punish these immoral behaviors. Through these strategies, most users and service providers avoid such spam behavior and obey the law. The servers that are in the "developing" stage will revert to the "secure and underlying" stage. Otherwise, the server may move to an even worse stage, the critical stage.

Stage 3: Critical. With ultra-weak *resilience* or no *resilience*, the server stays in the critical stage. Here, immoral behavior and the use of immoral technology may reach the pandemic level, and service providers cannot control their service quality. Internet governance cannot control these conditions.

 ISBN: 978-1-61208-390-2

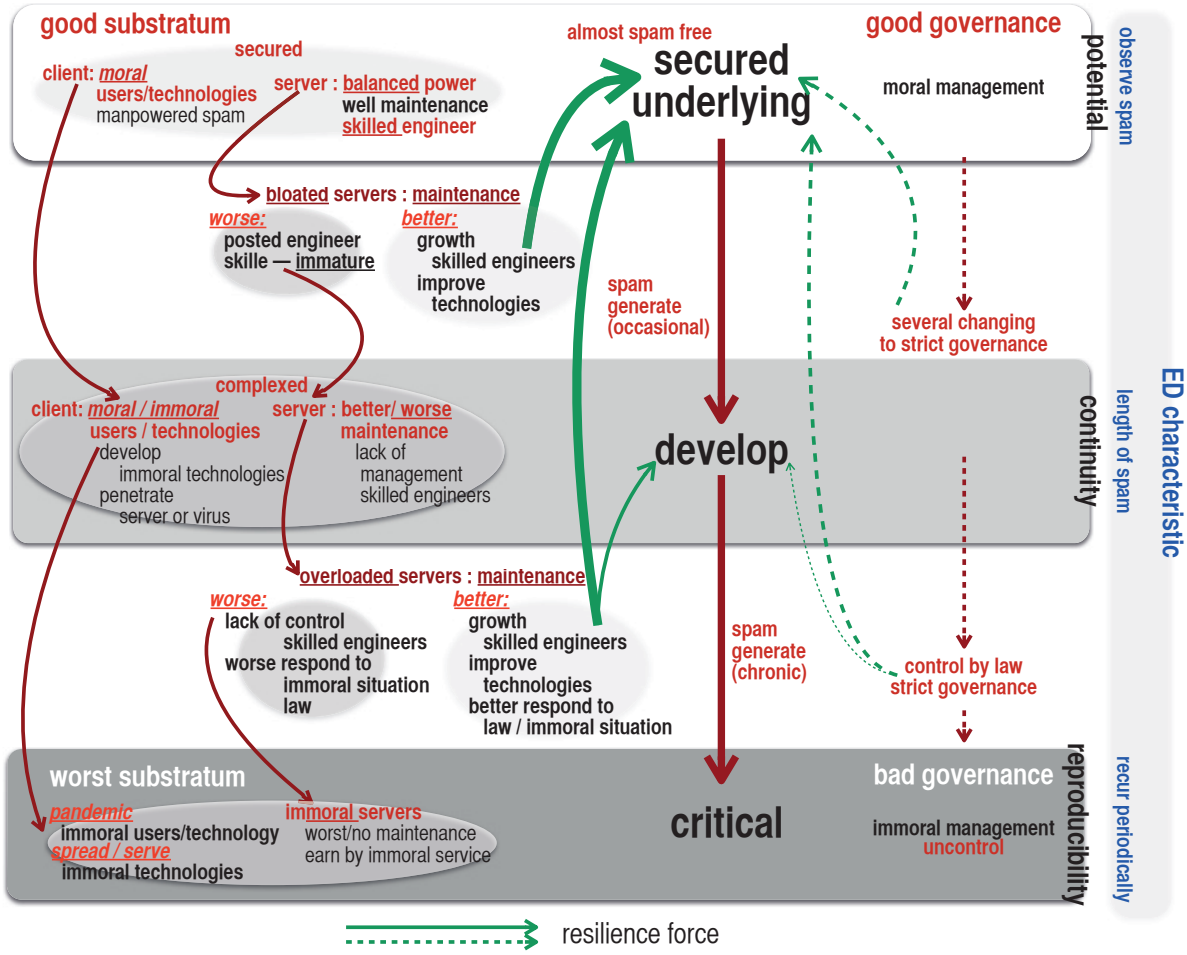


Figure 1. Microstructure for resilience.

III. Evolution Diagram and Its Derivatives for Measuring Strengths of Resilience Force

A. Evolution Diagram (ED)

An Evolution Diagram (ED) is a method of quantitatively representing the pattern of evolution history concerning a particular event (e.g., sending spam mail). For a given duration of observation of an event, (e.g., 1 year), we obtain a series of event-occurrence times as the data. We are interested in the tendency of change in the density of event occurrences over time, which should be related to the *resilience* force of the agent, (i.e., the spam server) that generates the event. An effective method of characterizing it is to analyze the data with different time resolutions (e.g., 1 year, 1/2 year, 1/4 year, and 1/8 year) and record "observed" if there is an event occurs in the time range or "not-observed" if there is none. Let the minimum observation time for detecting spam mail be τ (typically τ = 1sec), and the total observation time be T, which is a multiple of τ. By assigning "1" if an event is observed and "0" if none is observed, we can generate a series of 1s and 0s with the length of T/τ, denoted as $\vec{a} = (a_1, \cdots, a_{T/\tau})$.

In order to analyze the event-occurrence series in a variety of time resolutions with an arbitrary observation time unit, $\delta t = j \times \tau$, where $j = 1, \cdots, T/\tau$, we create $\vec{b}^{(j)} = (b_1, \cdots, b_i, \cdots, b_{L^{(j)}})$ from $\vec{a}$, where $L^{(j)}$ is the partition number calculated by the following formula:

$$L^{(j)} = \lfloor \frac{T}{j \times \tau} \rfloor. \quad (1)$$

In here, any positive integers in the range are allowed, but we use $j = 1, 2, 4, \cdots, 2^n (= T)$ for the sake of brevity of calculation.

The most coarse observation corresponds to $L^{T/\tau} = 1$, and the most precise observation corresponds to $L^{(1)} = T/\tau$. Here, $\vec{b}^{(j)}$ represents a seriese of event occurrences for the given time resolution $\delta t = j \times \tau$, as a series of 1s and 0s; $b_j = 1$ if the event is observed during the i-th observation period with the time resolution of δt, or 0 if non is observed. The ED value, $I(j)$, is calculated from $\vec{b}^{(j)}$ by applying the following formula:

$$I(j) = \sum_{i=1}^{L^{(j)}} b_i^{(j)} / L^{(j)}, \quad (2)$$

where $I(j)$ has value between 0 and 1.

An ED is created by plotting ED values as a function of j, where $j = 1, \cdots, T/\tau$, which is actually used to calculate

 ISBN: 978-1-61208-390-2

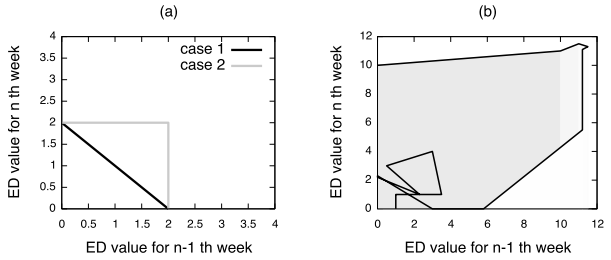


Figure 2. (a) Example of ED transition. (b) Example of calculation of Continuity and Reproducibility.

TABLE II. ED-VALUES IN THE CASE OF $T = 1$, TOTAL NUMBERS OF SPAMS ARE IN THE PARENTHESIS

Number of spams per 1 month		1 spam	5 spams	10 spams	30 spams
1 month		2.00 (1)	2.25 (5)	2.34 (10)	2.48 (30)
continue	2 months	2.13 (2)	2.99 (10)	3.17 (20)	3.45 (60)
	3 months	2.63 (3)	3.41 (15)	3.70 (30)	4.10 (60)
	4 months	2.88 (4)	3.72 (20)	4.06 (40)	5.57 (120)
equal interval	6 months	3.00 (2)	3.99 (10)	4.16 (20)	4.42 (60)
	4 months	3.50 (3)	4.37 (15)	4.64 (30)	5.06 (90)
	3 months	4.00 (4)	4.99 (20)	5.57 (40)	6.34 (120)

the ED values. ED values increase monotonically to 1 as j approaches its maximum value, T/τ, if if the server has sent at least one spam mail during the observation period.

The total ED value for the k-th observation series is calculated by summing the ED values for the actually taken j values:

$$\mathrm{ED}(k) = \sum_{j=\{1,\cdots,T/\tau\}} I(j). \tag{3}$$

B. Relationship between total ED values and Spam Sending Patterns: Simulation

Table II presents the results of the calculation of total ED values for artificially generated spam-sending patterns. The length of observation T is set to $T = 2^{24}$ sec, which is 194 days or 6.5 months, and $j = 2^0, 2^1, 2^2, 2^3, \cdots, 2^{24}$. The columns represent four different spam-sending patterns in terms of the total number of spams (1, 5, 10 and 30). It is assumed that spam-sending events occur periodically. For example, when the total number of spams is 10, one spam is sent every 3 days. The rows indicate the length of seven spam-sending patterns: 2, 3 or 6 months, and recurring patterns with 6-, 4-, and 3-month intervals. The figures in parentheses indicate the total number of spams sent in a year in the 28 different spam-sending conditions, for referencing purposes.

The total ED values are not necessarily proportional to the total number of spams. Comparison of the four cases where the total number of spams per year is 20 indicates that "5 spams/month with the equal interval of 3 months (i.e., 5, 0, 0, 5, 0, 0, $\cdots$)" has the worst total ED value (4.99). The case "10 spams/month lasting 2 months (i.e., 10, 10, 0, 0, 0, 0, $\cdots$)" has the lowest total ED value (3.17). The cases "5 spams/month lasting 4 months (i.e., 5, 5, 5, 5, 0, 0, $\cdots$)" with the total ED value of 3.72 and "10 spams/month with the equal interval of 6 months (i.e., 10, 0, 0, 0, 0, 10, 0, $\cdots$)" with the total ED value of 4.16 are in-between.

IV. RESILIENCE AND ED TRANSITION DIAGRAM

A. ED Transition Diagram

Introducing *resilience*, we can mark spam servers' states or degrees of healthiness. We derive *resilience* using an ED transition diagram as explained below. An ED transition diagram is defined on an $x-y$ plane. Each point has $(n-1)$-th total ED value as the x value and n-th total ED value as the y value. Combining the points for $n = 1, \cdots$, we can draw a figure with connected lines. In the following example, we calculate total ED values by setting $T = 2^{20}$, approximately 12 days. If $(n-1)$-th and n-th week's total ED values are different, we judge that there have been some changes in conditions.

Figure 2 (a) presents ED transition diagrams for two ideal cases. N is set to 20, with 1 week of observation time. In case 1's simulation, spam is detected only once every 6 months; the ED value is 2 when spam is detected and 0 when no spam is detected. Therefore, the coordinates change in the order (0, 0) $\rightarrow$ (0, 2) $\rightarrow$ (2, 0), and a corresponding triangle appears. In case 2's simulation, spam is detected every week. The coordinates change in the order (0, 0) $\rightarrow$ (0, 2) $\rightarrow$ (2, 2) $\rightarrow$ (2, 0), and a corresponding square appears. These two examples demonstrate that if the same ED value is calculated from different event occurrence patterns, different closed trajectories are obtained. Using these features, we define three elements for *resilience*: maximum transmission potential, continuity, and reproducibility.

Maximum transmission potential. The maximum transmission potential of *resilience* is determined from the area of the outer edge in the ED transition diagram. It is calculated by the area of a closed surface. In the example depicted in Fig. 2 (b), the gray zone is the value of maximum transmission potential. The outer edge of the ED transition diagram indicates the most active spam transmitting in the period. The area depends on the amount of spam transmission and the frequency for the most malicious condition of the server.

Continuity. Continuity represents the line integral of the ED transition diagram's trajectory. Practically, we calculate the Euclidean distances along the trajectory. The total distance depends on the duration of spam-sending.

Reproducibility. Lastly, we characterize spam transmission patterns by taking into account the direction of line segments that constitute the ED transition diagram. Here, we define six codes for y_n as week n's ED value:

- `S` : $y_{n-1} = 0,\ y_n = a$,
- `G` : $y_{n-1} = a,\ y_n = 0$,
- `_` : $y_{n-1} = a,\ y_n = a$,
- `E` : $y_{n-1} = 0,\ y_n = 0$ (repetition of `E` is counted as a single `E`),
- `U` : $y_{n-1} = a_1,\ y_n = a_2\ (a_1 < a_2)$,
- `D` : $y_{n-1} = a_2,\ y_n = a_1$.

For example, Fig. 2 (a) represents as `ESGE`, and Fig. 2 (b) represents as `ES_UGS`$\cdots$`GE`. In cases, Fig. 2 (a) has single `S` and (b) has several `S`, which means condition (b) is worse index value of reproducibility. Following these codes, reproducibility can be defined as the number of `S`.

 ISBN: 978-1-61208-390-2

TABLE III. CLASSIFICATION OF RESILIENCE FORCE, CHECKS ARE GIVEN FOR MALICIOUS COMPONENTS

class No.	reproducibility (R)	continuity (C)	potential (P)
0			
1(P)			√
2(C)		√	
3(CP)		√	√
4(R)	√		
5(RP)	√		√
6(RC)	√	√	
7(RCP)	√	√	√

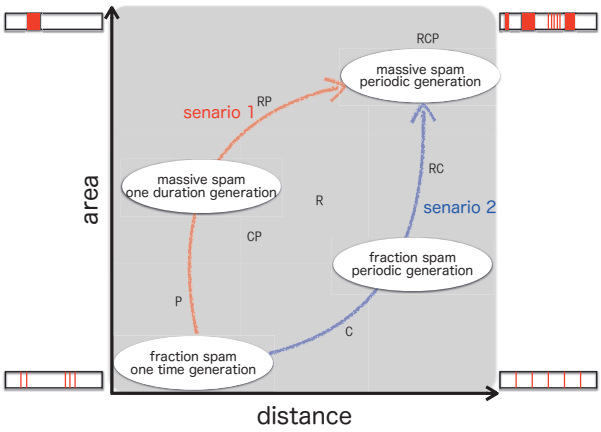


Figure 3. Relation of 2 senarios for spam server ecology and *resilience* class.

B. Resilience classes

Finally, we determine *resilience* for spam servers based on three components: maximum transmission potential, continuity, and reproducibility. We set the threshold value of the components to the values of the top 75% of these elements, and estimated the normalized degree of *resilience* of each server. Here, the normalization factor was the value for a server that sent spam mail once during the period.

Table III lists the *resilience* classes defined by counting the components whose values exceed the threshold values.

Most servers are classified as class 0. The ratios of No. 6 (CR), No. 1 (P), and No. 7 (PCR) are higher. Hence, No. 6 (RC) is server to send a long period of time a small amount of spam, No. 1 (P) is server to send a short period of time a large amount of spam, No. 7 (RCP) is server to send a long period of time a large amount of spam elements of the *resilience* has a meaning separate. Each component of *resilience* has a different meaning. Reproducibility R and continuity C mean weakness of *resilience*. Reproducibility is a measure of server status change and turbulence of ED value, and continuity is strongly related to the reproducibility cause of generating a line integral of ED transition. Hence, higher reproducibility and continuity indicate weak *resilience*. Maximum transmission potential is a useful component for measuring the strength of *resilience* in contrast with continuity and reproducibility. Even though maximum transmission potential is large whereas continuity and reproducibility are low, better control and management of servers derive low continuity and reproducibility. In a sense, maximum transmission potential functions with the combination of continuity and reproducibility.

V. TWO ECOLOGICAL SCENARIOS FOR SPAM SERVER

Using these analyses and simulations, we develop the "ecology of a spam server." Evans [8] summarized two distinct cognitive systems underlying reasoning: system 1, which is rapid, parallel, and automatic in nature; and system 2, which involves abstract hypothetical thinking. Coordinating these two systems, a human can make decisions. Applying this idea, we construct two ecological scenarios for a spam server (Fig. 3). First, we set two parameters, distance and area, in the ED transition diagram. Distance represents continuity with periodic spam-sending from the server. When a spam server tries to set the duration of a period, the server uses reasoning, such as safety (undetected by the administrator) duration. In this case, the distance is long but the area is small, as denoted by the line indicated as scenario 2 in Fig. 3. Area represents the maximum transmitting potential of the server. When a spam server tries to send many spam mails, it does not need to think about whether or not it has the potential to send so much mail. The only behavior of the server or spammer is sending spam without setting the duration of the period; it simply sends while the administrator does not stop the behavior. This behavior is denoted by the scenario 1 line in Fig. 3. Usually, System 2 monitors how System 1 behaves and warns System 1 when it captures System 1 attempts to send a lot of spam mails. Compared with Fig. 1, this stage is regarded as "secure but underlying." If system 1 or 2 arises in the server, the stage of the server will change to "developing." In this stage, if the *resilience* force is very strong, the server recovers its two systems' cooperation (i.e., the stage will return to "secure but underlying") . If the *resilience* force is weak, the server's two systems exhibit worse resonance: System 2 fails to warn System 1 not to send spam mails and as a consequence System 1 attempts to send spams more frequently in the shorter duration. The stage of the server will then be "critical" (upper right-hand area of Fig. 3). The stage of the server will then be "critical" (upper right-hand area of Fig. 3). Many resources are available to prevent changing to worse stages. One is security technology. In addition, human ability (e.g., administrators' management skills and users' morality) is also important. If human behavior does not change, the environment (e.g., Internet governance or regional management laws) must become stricter. Using analysis of spam servers' ecological features in the network's geological regions, we encourage regional management of the Internet. Finally, we observe features of the regional ecology of spam servers to provide several suggestions for management.

VI. OBSERVED ECOLOGY OF SPAM SERVERS

We collected "spam mail headers," which are the headers of mail identified as spam by mail-filtering software (SpamAssassin) at our university. Observation was conducted from March 1, 2013, through February 28, 2014. We identified 1,733,929 domains and 21,332,168 spam headers. For demonstration purposes, we randomly extracted 10,000 domains. The total ED value of each domain was calculated by setting $T = 1$ year and $j = 2^0, 2^1, 2^2, 2^3, \cdots, 2^{24}$. Note that the possible total ED values are limited. Only 5,703 domains have a total ED value of 2. As indicated in Table II, this value corresponds to the spam-sending pattern "one spam mail sent in a year." In addition, 2,109 domains have a total ED value of 2.5. Figure 4 plots the relationship between maximum transmitting potential and continuity for 10,000 spam domains, which are

 ISBN: 978-1-61208-390-2

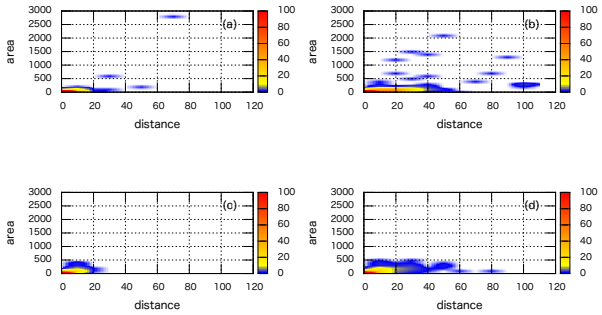


Figure 5. Sample of several stage at senario 1 and 2 for several ccTLD area. (a) senario 1 stage, (b) worst stage, (c) initial stage, (d) senario 2 stage.

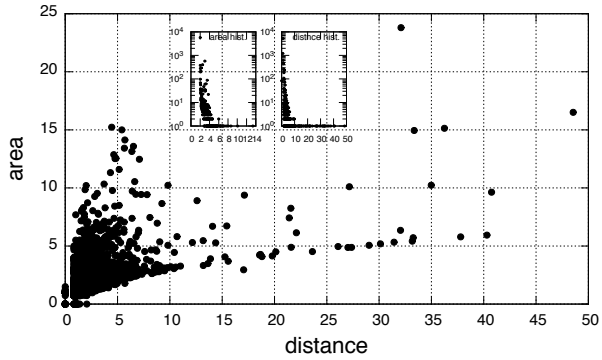


Figure 4. Relation of value of distance and area.

measured by the amount of outer area and line integral of ED transition diagram, respectively. for 10,000 spam domains. The graph depicts the distribution of each server's area (left side)/distance (right side). We find that the distribution of distancearea relationships exists between the four points in Fig. 3. Typically, large area and short distance are a healthy state for the transmission of spam in a short term (1 to 3 weeks). For this reason, the total ED value tends to become smaller. This class is equivalent to sending 30 spams in a month in Table II. Separating observed points in Fig. 4 from each geological server, we can determine the regional features of server management. In the graph, we show the histgram of area (left side)/distance (right side) value. Histgrams' horizontal axis represents normalized value of area/distance, and vertical axis represents the fraction of occurence. Figures 5 (a) through (d) present observed typical examples of the four geological regions' ecology of spam servers. The contour indicates the number of spam servers at the different values. We found four types of ecology derived from Fig. 3. For example, we regard (c) in Fig. 5 as "stable but underlying," (a) and (d) as "developing", and (b) as "critical." With the four features and microstructure in Fig. 1, we can estimate the network region's server management. Fig. 5 (a) represents that a number of servers are in the state of high continuity. In contrast, Fig. 5 (d) represents many servers are in the state of high reproducibility. Fig. 5 (b) shows a number of servers are in the state of high continuity or reproducibility. We can find these features in each ccTLD area. With the results, we can guess how the geological area's Regional/National Internet Registry makes Internet management policy: each Internet Registry has already had effective strategies for the Internet communication with specific reasons, or requires to enforce the area's Internet security of management.

VII. Conclusion

In this study, we proposed the concept "ecology of a spam server," based on the *e*-Network and *resilience* in an effort to understand the mechanism. First we considered the microstructure for *resilience* with the *e*-Network framework and demonstrated three stages for spam servers: secure but underlying, developing, and critical. Each stage included several features of potential/observe, continuity/length of spam, and reproducibility/periodic recurrence. Using these features, we introduced ED, a method for quantitatively representing the pattern of evolution history concerning a particular event, analyzing the data with different time resolutions and recording observed events. Using the ED values, we generated an ED transition diagram and defined three components to measure *resilience*: maximum transmission potential (calculated by area of closed surface), continuity (line integral for the ED transition diagram's trajectory), and reproducibility (number of restarts of spam emailing). Through these processes, we define eight classes of *resilience* with these three components and analyze tentative features of spam server behavior. Thus, we can advance the study of the ecology of a spam server, which will be a useful tool for producing strategies not only for security but also for the Internet/country domain governance.

Acknowledgment

The study described in this paper has been partially funded by the Scientific Research Expense Foundation C Representative: Katsuko T. Nakahira (24500308).

References

[1] K. T. Nakahira, "A Framework for Understanding Human e-Network – Interactions among Language, Governance, and more. " [Online]. Available: http://www.maayajo.org/IMG/SIMC/paris-v2.pdf[accessed2015-02-10]

[2] P. Graham, "Different Methods of Stopping Spam." [Online]. Available: http://www.windowsecurity.com/whitepapers/anti_spam/Stopping_Spam.html[accessed2014-09-10]

[3] F. Li and M. han Hsieh, "An Empirical Study of Clustering Behavior of Spammers and Groupbased Anti-Spam Strategies," in CEAS 2006 Third Conference on Email and AntiSpam, 2006, pp. 27–28.

[4] S. Stanković and D. Simič, "Defense Strategies Against Modern Botnets," CoRR, vol. abs/0906.3768, 2009. [Online]. Available: http://arxiv.org/abs/0906.3768

[5] F. V. Staden and H. S. Venter, "The State of the Art of Spam and Anti-Spam Strategies and a Possible Solution using Digital Forensics." in ISSA, H. S. Venter, M. Coetzee, and L. Labuschagne, Eds. ISSA, Pretoria, South Africa, 2009, pp. 437–454. [Online]. Available: http://www.bibsonomy.org/bibtex/2a914634f9302a8e9b54425ab149e0e5d/dblp[accessed2015-02-10]

[6] P. Smith, D. Hutchison, J. P. Sterbenz, M. Schöller, A. Fessi, M. Karaliopoulos, C. Lac, and B. Plattner, "Network Resilience: A Systematic Approach," IEEE Communications Magazine, vol. 49, no. 7, July 2011, pp. 88–97.

[7] "Measurement Frameworks and Metrics for Resilient Networks and Services: Technical Report," European Network and Information Security Agency, Tech. Rep., Feb. 2011.

[8] J. S. B. T. Evans, "In two minds: dual-process accounts of reasoning," Trends in Cognitive Sciences, vol. 7, no. 10, 2014/11/15, pp. 454–459. [Online]. Available: http://www.cell.com/trends/cognitive-sciences/abstract/S1364-6613(03)00225-0[accessed2015-02-10]

 ISBN: 978-1-61208-390-2

Simulation of the Emergence of Language Groups Using the Iterated Learning Model on Social Networks

Makoto Nakamura
Japan Legal Information Institute,
Graduate School of Law,
Nagoya University
Email: mnakamur@nagoya-u.jp

Ryuichi Matoba
Department of Electronics
and Computer Engineering,
National Institute of Technology,
Toyama College,
Email: rmatoba@nc-toyama.ac.jp

Satoshi Tojo
School of Information Science,
JAIST,
Email: tojo@jaist.ac.jp

***Abstract*—In evolutionary linguistics, the Iterated Learning Model (ILM) is often used for simulating the first language acquisition. Our purpose in this paper is to develop an agent-based model for language contact based on ILM. We put a learning agent on each node in the social network. Our experimental result showed that the language exposure rather deteriorates the emergence of local common languages, and grammars become non-compositional, which is different from our expectation. However, we have shown that an excessive string-clipping as well as a language exposure may constrain the appearance of local language community, independent of the shape of networks.**

Keywords–Simulation, Language Acquisition, Iterated Learning Model, Social Network.

I. INTRODUCTION

Thus far, simulation studies have played an important role in the field of the evolution of language [1]. Especially, a very important function of simulation is to prove if a prediction actually and consistently derives from a theory [2]. So far, there have been a variety of methodologies proposed on simulating the evolution of languages, each of which belongs to a different level of abstraction. Simulation studies for population dynamics alone include an agent-based model of language acquisition by Briscoe [3], which was developed toward a formal model of language acquisition device. On the other hand, Nowak [4] proposed a mathematical theory of the evolutionary dynamics of language called the language dynamics equation. The language dynamics equation is highly abstract, while agent-based model is considered to be a concrete, or less abstract.

Our goal is to provide a framework that represents the diachronic change in language by the contact among language communities. It would be useful not only for simulating typical language changes but also for novel phenomena taking place in the cyber world. In recent decades, the evolution of the Internet makes users possible to participate in discussions with anonymous people concerning their favorite topics beyond the physical distance. They do not only exchange small bits of information, but rather seem to establish a durable channel to communicate among people sharing common tastes on a chat or bulletin board system. They often employ spoken language instead of formal one after sharing common interests, and thus the expression tends to be spontaneous and haphazard. This phenomenon is often seen in language contact, but the time and size of the language change on the internet are extremely fast and large, respectively [5]. Using the framework, it would be possible to deal with this rapid language change as a phenomenon of language evolution.

There have been simulation models dealing with language change. We employ them as possible. Thus far, Nakamura et al. [7] proposed a mathematical framework for the emergence of creoles [6] based on the language dynamics equation. Toward more concrete analysis, they introduced a spatial structure to the mathematical framework [8] [9], in which learning agents contact with neighbors according to the learning algorithm. Furthermore, the spatial structure was expanded into complex networks [10]. Their studies are based on a hypothesis about the emergence of creoles, that is, language contact is likely to stimulate creolization. However, their learning mechanisms are too simple to observe language changes from a linguistic aspect, as languages are defined as similarity measures in a numeric matrix.

We propose an agent based model to deal with grammatical changes in the language community. Therefore, our purpose in this paper is to show a relationship between communication among learning agents and grammatical changes. We employ Simon Kirby's Iterated Learning Model (ILM) [11], which shows a process of grammatical evolution through generations. Kirby's ILM has often been used in simulation models concerning language evolution [12]. One important reason for this is that ILM is robust against input sentences in terms of a syntactic learning. As long as learning from a single parent, its infant agent receives sentences derived from a consistent grammar, it is possible to acquire a concise grammar. Currently, the learning situation in ILM is extended to a multiple families connecting with a network. We can observe the language change, not only in diachronic situation, i.e., in parent-child relation, but also in synchronic situation.

Thus far, we have shown a pilot version [13], where we found a problem reported by Smith and Hurford [20], that is, in the case learning agents potentially have more than one teacher agent, the length of syntax rules tends to increase rapidly over generations due to the addition of symbols of meaningless terminal symbols. This problem causes an unnatural learning, which results in a fatal combinatorial explosion. We solved this problem and try again with a larger number of agents.

 ISBN: 978-1-61208-390-2

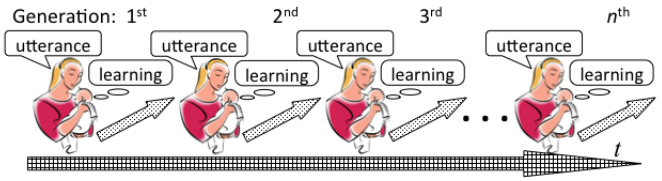


Figure 1. The iterated learning model

Verb: admire, detest, hate, like, love
Noun: john, mary, pete, heather, gavin

e.g.) *love(mary, john)*
(Identical arguments are prohibited.)

Figure 2. Words used in the experiment.

This paper is organized as follows. In Section 2, we introduce Kirby's ILM. In Section 3, we propose an agent-based model for language contact. In Section 4, we examine our proposed method, and conclude in Section 5.

II. Iterated Learning Model for Social Networks

In this section, we mention how to deal with ILM on the social networks. Firstly, we briefly explain Kirby's ILM. After that, we introduce the modification for social networks by Matoba et al. [14] in order to avoid the combinatorial explosion, which enables us the expansion of ILM.

A. Briefing Kirby's Iterated Learning Model

Kirby [11] introduced the notions of compositionality and recursion as fundamental features of grammar, and showed that they made a human possible to acquire compositional language. Figure 1 illustrates ILM. In each generation, an infant can acquire grammar in his/her mind given sample sentences from his/her mother. When the infant has grown up, he/she becomes the next parents to speak to a newborn baby with his/her grammar. As a result, infants can develop more compositional grammar through the generations. Note that the model focuses on the grammar change in multiple generations, not on that in one generation. Although the poverty of stimulus explains the necessity of the universal grammar [15], Kirby [11] modeled it as learning through bottlenecks, which are rather necessary for the learning. Also, he adopted the idea of two different domains of language [16]–[19], namely, I-language and E-language; I-language is the internal language corresponding to speaker's intention or meaning, while E-language is the external language, that is, utterances. In his model, a parent is a speaker agent and his/her infant is a listener agent. The speaker agent gives the listener agent a pair of a string of symbols as an utterance (E-language), and a predicate-argument structure (PAS) as its meaning (I-language). A number of utterances would form compositional grammar rules in listener's mind, through learning process. This process is iterated generation by generation, and converges to a compact, limited number of grammar rules.

According to Kirby's ILM, the parent agent gives the infant agent a pair of a string of symbols as an utterance, and PAS as its meaning. The agent's linguistic knowledge is a set of a pair of a meaning and a string of symbols, as follows.

$$S/love(john, mary) \rightarrow \text{hjsbs}, \tag{1}$$

where the meaning, that is the speaker's intention, is represented by a PAS *love(john, mary)* and the string of symbols is the utterance "hjsbs"; the symbol *'S'* stands for the category Sentence. The following rules can also generate the same utterance.

$$\begin{array}{rcl} S/love(x, mary) & \rightarrow & \text{h } N/x \text{ sbs} \\ N/john & \rightarrow & \text{j} \end{array} \tag{2}$$

where the variable x can be substituted for an arbitrary element of category *N*.

The infant agent has the ability to generalize his/her knowledge with learning. This generalizing process consists of the following three operations [11]; *chunk*, *merge*, and *replace*.

Chunk This operation takes pairs of rules and looks for the most-specific generalization.

$$\begin{array}{l} \left\{ \begin{array}{rcl} S/love(john, pete) & \rightarrow & \text{ivnre} \\ S/love(mary, pete) & \rightarrow & \text{ivnho} \end{array} \right. \\ \Rightarrow \left\{ \begin{array}{rcl} S/love(x, pete) & \rightarrow & \text{ivn } N/x \\ N/john & \rightarrow & \text{re} \\ N/mary & \rightarrow & \text{ho} \end{array} \right. \end{array} \tag{3}$$

Merge If two rules have the same meanings and strings, replace their nonterminal symbols with one common symbol.

Replace If a rule can be embedded in another rule, replace the terminal substrings with a compositional rule.

In Kirby's experiment [11], five predicates and five object words shown in Figure 2 are employed. Also, two identical arguments in a predicate like *love(john, john)* are prohibited. Thus, there are 100 distinct meanings (5 predicates $\times$ 5 possible first arguments $\times$ 4 possible second arguments) in a meaning space.

The key issue in ILM is to make the situation of *poverty of stimulus*. As long as an infant agent is given all sentences in the meaning space during learning, he/she does not need to make a compositional grammar; he/she would just memorize all the meaning-sentence pairs. Therefore, agents are given a part of sentences in the whole meaning space. The total number of utterances the infant agent receives during learning is parameterized. Since the number of utterances is limited, the infant agent cannot learn the whole meaning space, the size of which is 100; thus, to obtain the whole meaning space, the infant agent has to generalize his/her own knowledge by self-learning, i.e., *chunk, merge,* and *replace*. The parent agent receives a meaning selected from the meaning space, and utters it using her own grammar rules. When the parent agent cannot utter because of lack of her grammar rules, she invents a new rule. This process is called *invention*. Even if the invention does not work to complement the parent agent's grammar rules to utter, she utters a randomly composed sentence.

 ISBN: 978-1-61208-390-2

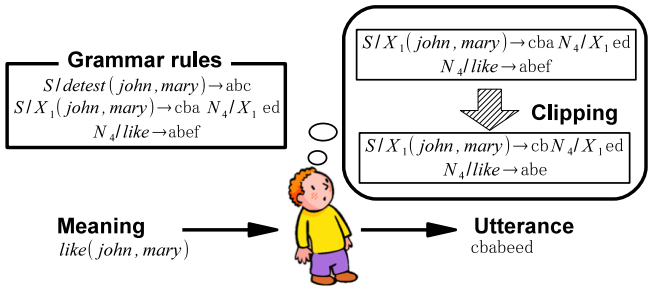


Figure 3. Image of clipping process

B. Process for String Clipping

When an infant agent has a number of teacher agents consisting of his/her parent and neighbors, as the teacher agents have their own compositional grammar rules, they are inconsistent with each other. Although the infant agent tries to find a common chunk among utterances, it would be a short string. Since there is little probability of making a chunk from short strings, only long ones are likely to survive toward next generations. As a result, learning agents tend to have compositional rules with extremely long strings over generations [20].

Matoba et al. [14] proposed a clipping process in their model, which solves the above problem. This process is called *backclipping*. After learning process of the infant agent, he/she curtails symbols in his/her grammar rules from the tail of string, unless it contains ambiguity. As a result, when the infant agent becomes the new parent agent in the next generation, the grammar set does not contain extremely long rules any more.

Figure 3 illustrates the clipping process in our model. The infant agent tries to utter strings of *like(john, mary)* as shortly as possible. Firstly, he/she chooses a grammar rule from his/her grammar set for generating utterance of *like(john, mary)*, and deletes symbols one by one, i.e., "cba", "ed", "abef". In case of "cba", this string does not exist in the grammar rules of the infant agent, then the infant agent executes backclipping, and the string becomes "cba" to "cb". The string "cb" does not exist in the grammar rules of the infant agent, so the infant agent executes backclipping, and the string becomes "cb" to "c". Since "c" exists in the grammar rules of infant agent, the infant agent does not abridge it anymore, and adopts "cb" as the clipped word of "cba". The same process is also applied to the other words. As a result, the sentence becomes shortened from "cbaabefed" to "cbabeed".

Actually, such phenomenon occurs in the real world, as cutting the beginning and/or the end of a word off. The deletion of a part of a word constructs a new and shorter word;

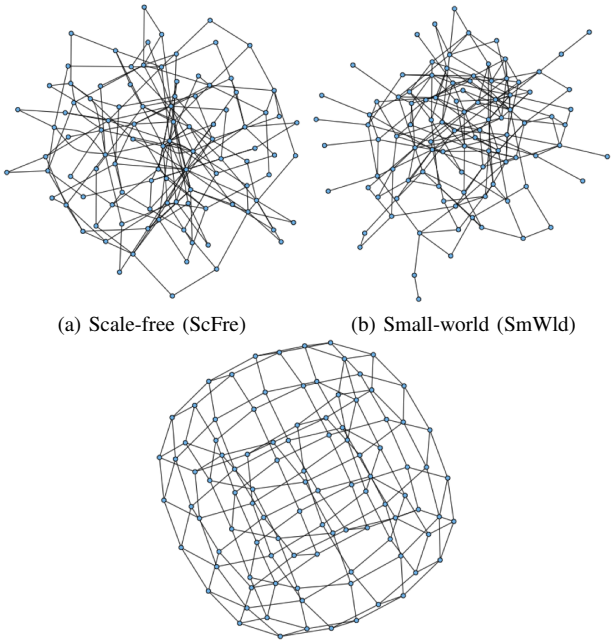
e.g.) Ham**burger** → **burger**, In**flu**enza → **flu**, **Exam**ination → **exam**

A position of clipping is dependent on a phonological reason [21]. Since ILM does not deal with phonological information, we need to find an alternative way to shorten strings. In English, we often omit a few syllables of each word [22];

e.g.) **ad**vertisement, **doc**tor, **lab**oratory, **prof**essor, **demo**nstration, **cap**tain, etc.

TABLE I. NETWORK CHARACTERISTICS

	Network type ($N = 100$)	Average Degree	Average shortest path
	Complete graph	99.00	1.00
	Star	1.98	1.98
(a)	Scale-free	3.96	3.28
(b)	Small-world	4.00	3.41
(c)	2D lattice	4.00	12.88
	Ring	2.00	25.25



(a) Scale-free (ScFre) (b) Small-world (SmWld)

(c) 2D-lattice (Lat2D)

Figure 4. Examples of the networks ($N = 100$)

III. Agent-based Model for Language Contact

In this section, we explain how language groups emerge in the agent-based model. Agents can get contact with neighbors on the network (Section III-A), who speak to the infant in a certain ratio of language exposure (Section III-B). The communication may affect agents' languages, which are classified into groups by the language similarity (Section III-C).

A. Social Networks for Language Communities

Social networks play an important role of language change, regardless of whether they are connected by an actual or virtual relationship. Some simulation studies deal with complex networks [10] [23] [24]. There are several types of networks, each of which characterizes many real-world communities.

Table I shows network characteristics, in which each value is calculated based on 100 nodes [24]. The average degree denotes the average number of edges connected to a node. The average shortest path length stands for the average smallest number of edges, via which any two nodes in the network can be connected to each other. In this paper, we examine scale-free, small-world and 2D-lattice networks.

Figure 4 shows examples of networks, in which the preceding two networks are regarded as complex networks and the latter is for comparison. Each agent is assigned on an node in the networks.

 ISBN: 978-1-61208-390-2

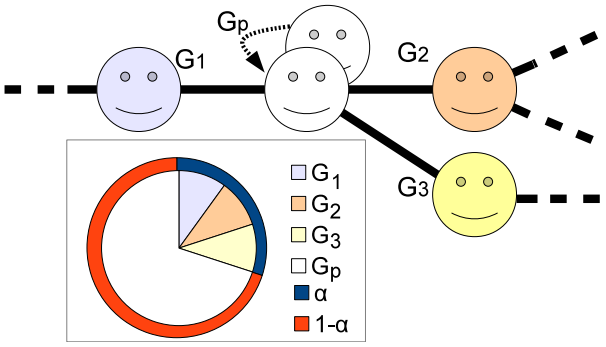


Figure 5. Language input from neighbors depending on the exposure ratio α

B. Exposure Ratio α

Nakamura et al. [7] has introduced an exposure ratio α, which determines how often language learners are exposed to a variety of language speakers other than their parents. They modified the learning algorithm of Nowak et al. [4], taking the exposure ratio into account in order to model the emergence of creole community. They have shown that a certain range of α is necessary for a creole to emerge. This parameter was further employed for the following network studies [8]–[10].

In some communities, a child learns language not only from his/her parents but also from other adults, whose language may be different from the parental one. In such a situation, the child is exposed to other languages, and thus may learn the most communicative language. In order to assess how often the child is exposed to other languages, let us divide the language input into two categories: one is from his/her parents, and the other is from other language speakers. The ratio of the latter to the total amount of language input is called an *exposure ratio* α. This α is subdivided into smaller ratios corresponding to those other languages, where each ratio is in proportion to the population of the language speakers.

An example distribution of languages is shown in Figure 5. Let G_i be the language of Agent i. Suppose a child has parents who speak G_p, he/she receives input sentences from G_p in the proportion of $1-\alpha$, and from non-parental languages $G_i (i \neq p)$ in the proportion of αx_i, where x_i denotes a population ratio of G_i speakers in the neighbors.

C. Distance between Languages and Language Groups

In this section, we discuss how to deal with languages in the framework of ILM on a social network. An infant receives a meaning-signal pair from his parent and neighbors according to the exposure ratio α. The number of utterances an infant receives is fixed, and he/she receives them in proportional to the language distribution for neighbors like the pie chart shown in Figure 5. The population consists of non-overlapping generations, that is, infants at each generation are born at the same time, become parents at the same time, and die at the same time. The network is fixed through generations.

In order to compare between languages, we define the distance in languages by the edit distance, known as the Levenshtein distance [25]; we count the number of insertion/elimination operations to change one word into the other. For example, the distance between "abc" and "bcd" becomes 2 (erase 'a' and insert 'd'). Once the learning process has been finished, each agent has his/her own grammar rules. In other words, each agent can enumerate all the sentences he/she can utter as E-language derived from I-language. Figure 6 depicts an image of enumeration. Note that all the compositional grammar rules are expanded into a set of *holistic rules,* which do not include any variable, i.e., a rule consists of a sequence of terminal symbols. Since the Levenshtein distance between corresponding strings can be calculated, the average distance normalized at the range from 0 to 1 comes into the distance between languages.

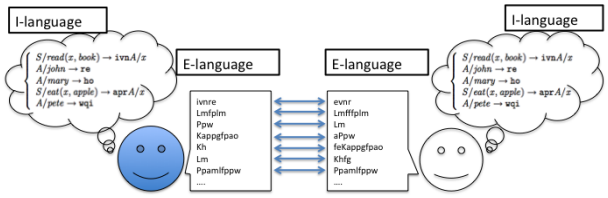


Figure 6. Calculation of the distance between languages

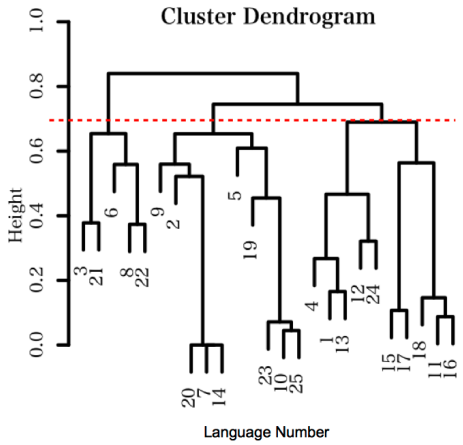


Figure 7. Clustering languages ($N = 25$)

Since agents independently invent languages, their acquired languages are different from each other. In order to classify agents into groups by the language similarity, we introduce a clustering method, recognizing a cluster as a language group. The relationship among languages is represented by a dendrogram shown in Figure 7. The vertical axis denotes the height of the tree, which generally depicts the mergers or divisions which have been made at successive level. We employed the complete linkage method throughout the experiments. The number of languages, therefore, depends on the cutting point of the tree. In this case, the community is regarded as consisting of three-language groups at the height of $\theta = 0.7$.

IV. Experimental Results

Our purpose of these experiments is to examine how the configuration of networks affects the language learning by infant agents. We expect language groups to emerge depending on the types of networks and other conditions. Therefore, we examine three types of networks; Scale-free, Small-world, and 2D-lattice networks. Scale-free and small-world networks are drawn with BA [26] and WS [27] models, respectively. The number of nodes is fixed to $N = 100$. In BA networks, the

 ISBN: 978-1-61208-390-2

number of edges to add in each step is 2, and the power is set to 0.2. The generation of multiple edges is allowed. In WS, the number of neighborhood is 2, and the rewriting probability is set to 1.

We measure (a) the number of *language groups*, (b) the number of *grammar rules* and (c) *expressivity* of the grammar. (a) is calculated by setting the threshold to distinguish languages in the dendrogram to $\theta = 0.7$. (b) denotes the average number of grammar rules created in an agent at Generation 100. (c) is defined as the ratio of the number of utterable meanings derived from the grammar rules to the whole meaning space. Each infant agent receives 50 sentences, while the meaning space is 100 (5 predicates $\times$ 5 possible first arguments $\times$ 4 possible second arguments). Therefore, agents need to acquire a compositional grammar for high expressivity.

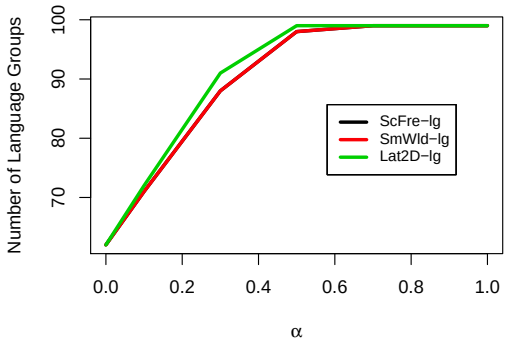
Since the exposure ratio α activates communication with neighbors, we also expect local dialects to emerge through communication. We parameterize $0 \leq \alpha \leq 1$, where the larger the value α is, the more frequently the neighbors speak to the infant agent. The situation $\alpha = 1$ is an extreme case that the infant's parent does not speak to him/her at all, but neighbors do.

Figure 8 shows experimental results. All the data are an average of 50 trials. Since scale-free and small-world networks are randomly drawn for each trial, no phenomenon peculiar to a specific network appears in the results.

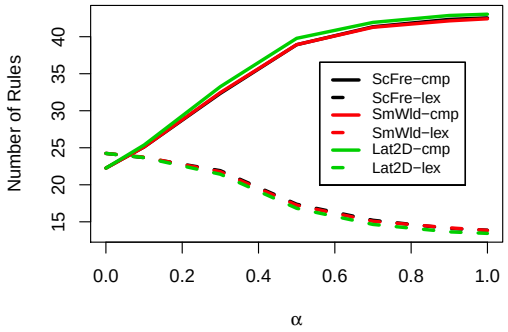
We classified languages into groups at 100th-Generation. Cutting the dendrogram at $\theta = 0.7$, we count the number of language groups in the community. Figure 8a shows the change in the number of language groups for every α. The labels "ScFre-lg," "SmWld-lg" and "Lat2D-lg" denote the numbers of language groups in the scale-free, small-world, and 2D-lattice networks, respectively.

Analyzing the results in Figure 8a, we can imply that each agent speaks a language different from others, as long as the number of language groups is almost the same as the population of the community. The language exposure is expected to make neighbors share a common language, but the result became different from our expectation. The reason is considered that the clipping process in ILM takes a chance for chunking from non-compositional sentences uttered by different agents. Finally, the most important thing is that there is no big difference between three networks. The exposure ratio α seems too effective to show minor difference between them. Another reason comes from the number of agents, which is insufficient to show difference between the networks. Despite employment of the clipping process, it was difficult to increase the number of agents more than 100 in a practical computational time.

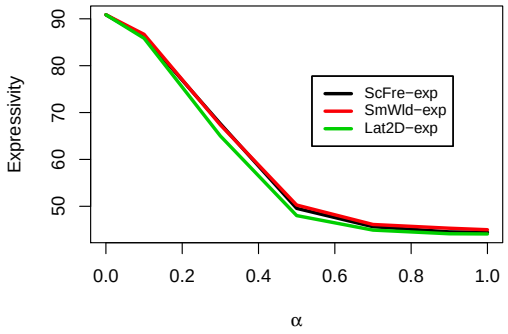
We also investigate acquired grammars. The number of grammar rules and its expressivity are shown in Figures 8b and 8c, respectively. Figure 8b shows that grammars become non-compositional according to α. The suffix '-cmp' denotes the number of compositional and holistic rules and the suffix '-lex' is the number of lexical rules. The decrease of lexical rules implies holistic rules occupy agents' knowledge, while an ideal compositional grammar consists of one compositional rule and ten lexical rules. Figure 8c is inevitably reflected by the compositionality. The language exposure negatively affects common languages. There is little difference between networks



(a) Language groups ($\theta = 0.7$)



(b) Grammar rules



(c) Expressivity (%)

Figure 8. Experimental results

in grammatical analysis as well as the result of language groups.

The series of experimental results differs from our expectation and from the former studies [7]–[10]. However, we have shown that an excessive string-clipping as well as a larger value of exposure ratio may constrain the appearance of local language community, independent of the shape of networks.

 ISBN: 978-1-61208-390-2

V. CONCLUSION

In this paper, we proposed an agent-based model for language contact. We employed Kirby's iterated learning model and complex networks. Languages are measured with the Levenshtein distance of utterances, which enables us to show the language divergence by the clustering. The language exposure is expected to make neighbors communicate with each other. Totally, we succeeded to implement a linguistic community with learning agents connected with a social network. The network model makes it possible to observe not only diachronic but also synchronic changes in grammar. We achieved implementation of a large-scale, agent-based model where 100 processes of ILM run in parallel, which contributes to the simulation study on language evolution.

Although we had been faced with a serious problem in terms of constructing a network model with ILM, the new method for a string clipping solved the combinatorial explosion. We need to investigate the algorithm of a string clipping and grasp why it works wrong for language contact.

In the near future, we plan to run more different types of simulations toward the framework of for the diachronic change in languages by language contact.

ACKNOWLEDGMENT

This work was partly supported by Grant-in-Aid for Young Scientists (B) (KAKENHI) No. 23700310, and Grant-in-Aid for Scientific Research (C) (KAKENHI) No.25330434 from MEXT Japan.

REFERENCES

[1] C. Lyon, C. Nehaniv, and A. Cangelosi, Eds., Emergence of Communication and Language. Springer, 2007.

[2] A. Cangelosi and D. Parisi, Eds., Simulating the Evolution of Language. London: Springer, 2002.

[3] E. J. Briscoe, "Grammatical acquisition and linguistic selection," in Linguistic Evolution through Language Acquisition: Formal and Computational Models, T. Briscoe, Ed. Cambridge University Press, 2002, ch. 9.

[4] M. A. Nowak, N. L. Komarova, and P. Niyogi, "Evolution of universal grammar," Science, vol. 291, 2001, pp. 114–118.

[5] D. Crystal, Internet Linguistics: A Student Guide, 1st ed. New York, NY, 10001: Routledge, 2011.

[6] J. Arends, P. Muysken, and N. Smith, Eds., Pidgins and Creoles. Amsterdam: John Benjamins Publishing Co., 1994.

[7] ——, "Exposure dependent creolization in language dynamics equation," in New Frontiers in Artificial Intelligence, ser. Lecture Notes in Artificial Intelligence, A. Sakurai, K. Hasida, and K. Nitta, Eds., vol. 3609. Springer, 2006, pp. 295–304.

[8] ——, "Self-organization of creole community in spatial language dynamics," in Proc. of 2nd IEEE International Conference on Self-Adaptive and Self-Organizing Systems (SASO2008), Venice, 2008, pp. 459–460.

[9] ——, "Prediction of creole emergence in spatial language dynamics," in LATA 2009 (Proc. of 3rd International Conference on Language and Automata Theory and Applications), ser. Lecture Notes in Artificial Intelligence, A.H.Dediu, A.M.Ionescu, and C.Martin-Vide, Eds., vol. 5457. Tarragona: Springer, 2009, pp. 614–625.

[10] ——, "Self-organization of creole community in a scale-free network," in Proc. of 3nd IEEE International Conference on Self-Adaptive and Self-Organizing Systems (SASO2009), San Francisco, 2009, pp. 293–294.

[11] S. Kirby, "Learning, bottlenecks and the evolution of recursive syntax," in Linguistic Evolution through Language Acquisition: Formal and Computational Models, T. Briscoe, Ed. Cambridge University Press, 2002.

[12] M. Delz, B. Layer, S. Schulz, and J. Wahle, "Overgeneralization of Verbs - the Change of the German Verb System," in Proc. of EVOLANG9, 2012, pp. 96–103.

[13] ——, "Multilayered formalisms for language contact," in Proc. of WS on Constructive Approaches to Language Evolution, Kyoto, 2012, pp. 145–147.

[14] R. Matoba, H. Sudo, M. Nakamura, S. Hagiwara, and S. Tojo, "Process Acceleration in the Iterated Learning Model with String Clipping," International Journal of Computer and Communication Engineering, vol. 4, no. 2, 2014.

[15] N. Chomsky, Rules and Representations. Oxford: Basil Blackwell, 1980.

[16] D. Bickerton, Language and Species. University of Chicago Press, 1990.

[17] N. Chomsky, Knowledge of Language:Its Nature, Origin, and Use. New York: Praeger, 1986.

[18] J. R. Hurford, Language and Number: the Emergence of a Cognitive System. Oxford: Basil Blackwell, 1987.

[19] S. Kirby, Function, Selection, and Innateness: The Emergence of Language Universals. Oxford University Press, 1999.

[20] K. Smith and J. R. Hurford, "Language Evolution in Populations: Extending the Iterated Learning Model," in Proc. of ECAL03, 2003, pp. 507–516.

[21] D. Jamet, "A Morphological Approach of Clipping in English. Can the Study of Clipping Be Formalized?" Lexis, no. 1, 2009, pp. 15–31.

[22] A. Veisbergs, "Clipping in English and Latvian," Poznan Studies in Contemporary Linguistics, no. 35, 1999, pp. 153–163.

[23] X. Castelló et al., "Modelling language competition: bilingualism and complex social networks," in Proc. of EVOLANG7, 2008, pp. 59–66.

[24] T. Gong, L. Shuai, M. Tamariz, and G. Jäger, "Studying Language Change Using Price Equation and Pólya-urn Dynamics." PLoS One, vol. 7, no. 3:e33171, 2012.

[25] D. Jurafsky and J. H. Martin, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2000.

[26] A.-L. Barabasi and R. Albert, "Emergence of scaling in random networks," Science, vol. 286, no. 5439, 1999, pp. 509–512.

[27] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," Nature, vol. 393, 1998, pp. 440–442.

 ISBN: 978-1-61208-390-2

Adaptive Anomalies Detection with Deep Network

Chao Wu, Yike Guo
Department of Computing
Imperial College London, London, UK, SW7 2AZ
Email: {chao.wu, y.guo}@imperial.ac.uk

Yajie Ma
College of Information Science and Engineering
Wuhan University of Science and Technology, Wuhan, China, 430081
Email: mayajie@wust.edu.cn

***Abstract*—In this paper, we try to apply inspirations from human cognition to design a more intelligent sensing and modeling system, which can adaptively detect anomalies. The target of intelligent sensing and modeling is not to get as much data as possible, or to build the most accurate model, but to establish an adaptive representation of sensing target and achieve balance between sensing performance requirement and system resource consumption. To achieve this goal, we adopt a working memory mechanism to facilitate the model to evolve with the target. We use a deep network with autoencoders as model representation, which is capable to model complex data with its nonlinear and hierarchical architecture. Since we typically only have partial observations from sensed target, we design a variance of autoencoder which can reconstruct corrupted input. We utilize attentional surprise mechanism to control model update. Training of the deep network is driven by surprises (which are also anomalies) detected (with data in working memory), which means model failure or target's new behavior. Due to partial observations, we are not able to minimize free-energy in a single round, but iteratively minimize it by keeping finding new optimization bounds. While both random and non-random sensor selection can create new optimization bounds, certain non-random methods like surprise minimization algorithm used in this paper demonstrate better performance. For evaluation, we conducted experiments on simulated data to test whether our methodology makes the model more adaptive, and got positive result. In the next step, we will try to apply the work on some real applications including ECG and EEG anomaly detection.**

Keywords–Cognitive sensing; deep learning; anomaly detection.

I. INTRODUCTION

The world is always dynamic, unpredictable, ambiguous, and noisy. Such uncertainty is the only reason why we need to be equipped with intelligence. From a cognitive point of view, the intelligence of any intelligent agent (like animals), is a kind of ability to achieve equilibrium with its uncertain environment. It can sense or act (to fit with or intervene its environment), to minimize its free energy [1]. Such intelligence is under some constraints. Facing dynamic and high-dimensional world, even for the most complex systems as human brains, neural computation resource is limited [4]. Also our action capabilities on the environment (like motion capability) is limited.

To handle the uncertain world with constrained resources, intelligent agent developed several crucial cognitive mechanisms: We use attention [6] and surprise [5] to select information deserved to be processed and allocate neural resource for feature binding; we have long-term / short-term working memory to organize the knowledge hierarchically; Our brain is constituted with a large amount of deep networks, each of which act as a universal model and support "one learning algorithm". Such deep networks are considered to be very useful to organize and process our knowledge. Whether these mechanisms, or at least some of them, could be adopted in sensing system to make it more intelligent? This idea motivated the work in this paper: we tried to get inspirations from biological intelligence, and design an adaptive computational framework for sensing and modeling a dynamic target, under system resource constrain. The emphasis here is to detect target's anomalies, or surprises, which indicates there appear some events or new behaviors of the target. The examples of anomalies include traffic accidents (*events*) on a road (*sensing target*), and seizures (*events*) in human brain (*sensing target*).

In section 2, some related works are discussed. Detailed methodology of our framework is given in Section 3. Section 4 provides both simulated evaluation and demonstration of an application. In Section 4, we conclude the paper with a future research plan.

II. RELATED WORKS

Historically, designing computing system by learning human cognition is not new. Cognitive science discoveries had inspired a lot of researches on unsupervised learning (e.g., the Analysis by Synthesis approach [7]). Another example is the formation of Infomax principle and independent component analysis [10], [11] from the inspiration of efficient coding [8], [9]. However, it's until recent years that the advance of cognitive science and neuroscience makes it possible to build a clearer picture of how our brain works. Based these findings, we believe the meeting with Brain Informatics (BI) [12], [13] and AI will be the next drive for both developments.

Cognitive perspective also inspired new insight of anomaly. We use an unsupervised method to train our model, and this model can reconstruct the input (with low loss function output) when the input is seen before (within the range of model's representation capability). Only once the observation is new to the model and cannot be reconstructed, we detect an anomaly. It is slightly different from ordinary anomaly detection, which excludes the outliers into the model, but related to novelty detection [16], which try to detect emergent and novel patterns in the data, and incorporated into the normal model after being

 ISBN: 978-1-61208-390-2

detected. This anomaly is also seen as an attentional surprise (more formally, 'surprisal'), induced by a mismatch between the sensory signals encountered and those predicted. Such surprise play the similar role of surprise for human, which acts as a kind of proxy for sensory information [14] (to select input for process). Similar principle has been applied in methods like predictive coding [15]

III. METHODS

A. Notations and problem definition

Target: We represent the target that we want to understand as $x = (x_1, x_2, ..., x_i, ...x_n) \in R^n$ with dimension n, which is the size of spatial-temporal resolution of the target (e.g. $n = 24$, if the target is the hourly temperature in one day). x can have infinite dimensions and $n = \infty$. x is a time series : at time t, we have $x^t = (x_1, x_2, ..., x_i, ...x_n)^t$, and we denote the collection of x as $X = \{x^1, x^2, ..., x^m\}$.

Observation: In most cases, we can not observe x' directly, but only observe partial x (in lower dimension k) with noise ϵ. We define this noisy and partial observations from sensors as $X' = \{x'^1, x'^2, ..., x'^m\}$, where $\|x'\|_0 = k$ and $k \leq n$. We assume $x' = s(x)_k + \epsilon$, where function $s(x)_k$ selects k elements in x, keeps their values, and sets the other elements to 0.

Model: Model y is established based on observations. It can be viewed as a representation of x, governed by some parameters θ, just like the internal model of human can be viewed as some higher representation of its input. y can be a distribution, a learned dictionary, or a stacked autoencoder as used in this paper.

Problem definition: For a sensing system, the goal is to learn $p(x)$, which is difficult when $p(x)$ is changing. So we try to approximate $p(x)$ with the mapping from y to x (with observations x'): $q(x|y; x')$, or simply $q(x|y)$. For this approximation, there are mainly two challenges: 1) mapping from y to x should not only approximate $p(x)$, but be able to adapt to the change of $p(x)$; 2) with only partial observations, we want them to be informative, so how to design function $s(x)_k$ is then crucial, and thus becomes well-known sensor selection problem. With this adaptive mapping from y to x, we then evaluate the new observations with the established model to check whether model fails. If so, we say there is **anomaly**.

B. Cognitive approach

According to free energy principle [1], any intelligent agent will sense or act to minimize its free energy $F(x, y)$, which is the KL divergence between its internal approximation with model and real environment (i.e. sensing target, in this context):

$$F(x, y) = D_{KL}(q(x|y; x')\|p(x)) \tag{1}$$

As a result, we minimize the KL divergence between these two:

$$D_{KL}(q(x|y)\|p(x)) = \int q(x|y) \ln \frac{q(x|y)}{p(x)} dx \tag{2}$$

$$= D_{KL}(q(x|y)\|p(x|y)) - \ln p(y) \tag{3}$$

Here we get two components: 1) the first component is the KL divergence between $q(x|y)$ and $p(x|y)$. $q(x|y)$ represents the approximated distribution of x given y as the internal model; and $p(x|y)$ represents the likelihood of x if it's governed by y. We can unbiasly estimate this component by replacing x with x', then $p(x'|y)$ becomes the empirical likelihood. In this paper, we define this component (or its approximation) as **surprise**. Once we have great surprise, it means the model is too rough (not well-trained), or the target exhibits some new behavior that has not been captured by the current model. In both situations, the model cannot capture the target, and needs to be updated. Cognition model will update y and thus reduce the KL divergence between $q(x|y)$ and $p(x|y)$. However, because we only use partial observation, we actually minimize a part of original KL divergence:

$$D_{KL}(q(x|y)\|p(x|y)) \leq D_{KL}(q(x'|y)\|p(x'|y)) \tag{4}$$

After we minimizing this KL divergence with x', a very important next step is to find a new bound of KL divergence optimization, by selecting new x' through random or non-random sensor selection, and then update the model in the next iteration; 2) the second component is the negative log of $p(y)$. It measures how unlikely a representation y will happened, and will be large if the sensing target is too dynamic or noisy. The optimization of this component is out of the scope of this paper.

We can also interpret this cognitive approach with Infomax principle. Best mapping from y to x maximizes the mutual information between x and y, which can be represented as the difference between entropy of x ($H(x)$) and conditional entropy of x given y ($H(x|y)$). Following the same approach in [2], we assume x comes from an unknown distribution $p(x)$ on which θ has no influence, so $H(x)$ is constant. Therefore, the target is then to maximize $-H(x|y)$, which by definition is:

$$\operatorname*{argmax}_y E_{p(x,y)}[\log p(x|y)] \tag{5}$$

With approximation $q(x|y)$, we have:

$$E_{p(x,y)}[\log q(x|y)] \leq E_{p(x,y)}[\log p(x|y)] \tag{6}$$

which is a lower bound of $-H(x|y)$. Assume we transform x to y with a deterministic or stochastic mapping $y = f_\theta(x)$(encoding), and reconstruct x from y by $x = g_{\theta'}(y)$ (decoding); and use empirical average over the observations as an unbiased estimation. We end up maximizing the mutual information in the following form:

$$\operatorname*{argmax}_{\theta,\theta'} E_{p(x)}[\log q(x|y = f_\theta(x); \theta')] \tag{7}$$

This corresponds to the reconstruction error criterion for autoencoders.

Before describing the details in our methodology, let's give an overall workflow of the methodology. We use some sensor selection algorithm $s(x)_k$ upon target x to get observation x', and use a working memory to store the data required for model training and update. Model training component trains

 ISBN: 978-1-61208-390-2

and updates y (its parameter θ) with data in memory. The established model is tested with new observation, to detect anomaly. Anomaly triggers actions including memory update and model update, and thus causes computational cost. Such model update optimize the free-energy within current bound, and sensor selection algorithm keeps trying to find new bound of free-energy for further optimization.

C. Model representation

We utilize a variation of stacked autoencoders [3] as the model representation. An autoencoder neural network is (unsupervisedly) trained with back propagation, setting the output values to be equal to the inputs. The result network takes an input x and transforms it to a hidden representation $y \in R^d$ through a deterministic function (with sigmoid activation function s):

$$y = f_\theta(x) = s(Wx + b) \tag{8}$$

It is parameterized by $\theta = \{W, b\}$. W is a $d \times n$ weight matrix. b is the bias vector. The resulting y is then mapped back to a reconstructed vector $z \in R^n$ in input space

$$z = g_{\theta'}(y) = g_{\theta'}(f_\theta(x^i)) = s(W'y + b') \tag{9}$$

with $\theta' = \{W', b'\}$. The weight matrix W' of the reverse mapping is constrained by $W' = W^T$. We get the optimized parameters θ^* and θ'^* by minimizing the reconstruction error (with loss function L) between x^i and z^i:

$$\theta^*, \theta'^* = \underset{\theta,\theta'}{\operatorname{argmin}} E_{p(x)}[L(x, z)] \tag{10}$$

Since:

$$L(x, z) \propto -\log p(x|z) \tag{11}$$

and thus

$$L(x, z) \propto -\log q(x|z) \tag{12}$$

We have:

$$\theta^*, \theta'^* = \underset{\theta,\theta'}{\operatorname{argmax}} \log q(x|z = g_{\theta'}(f_\theta(x))) \tag{13}$$

$$= \underset{\theta,\theta'}{\operatorname{argmax}} \log q(x|y = f_\theta(x), \theta') \tag{14}$$

which is the same form of optimization described before. With m training set, we will try to optimize:

$$= \underset{\theta,\theta'}{\operatorname{argmax}} \frac{1}{m} \sum_{i=1}^{m} L(x^i, g_{\theta'}(f_\theta(x^i))) \tag{15}$$

By placing constraints on the network, we can discover structure about the data and learn useful representation. Instead of limiting the size of hidden layer ($\|y\|_0$), we allow the size to be large, but impose sparsity constraints. An extra penalty term is added to optimization objective. let $(\hat{\rho})_j$ be the average activation of hidden unit j, averaged over the m training examples:

$$\hat{\rho}_j = \frac{1}{m} \sum_{i=1}^{m} [a_j(x^i)] \tag{16}$$

where a_j denotes the activation of this hidden unit when the network is given a specific input x. And let ρ to be a sparsity parameter, typically a small value close to zero (e.g. 0.05). So the penalty term is KL divergence between two (s_l is size of l-th hidden layer):

$$\sum_{j=1}^{s_l} KL(\rho \| \hat{\rho}_j) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \frac{1 - \rho}{1 - \hat{\rho}_j} \tag{17}$$

Such autoencoder is used as a building block to train deep networks, with the learned representation of the k-th layer used as input for the $(k+1)$-th to learn a second level representation. Therefore, we train the $(k + 1)$-th layer after the k-th has been trained. These layers are "stacked" in a greedy layer-wise approach as deep RBMs. This greedy layer-wise procedure has been shown to yield significantly better performance than random initialization.

D. Memory

Inspired by human short-term working memory, we enable a simple memory for the data modeling: $M = \{x^{*1}, x^{*2}, ..., x^{*k}\}$. k is the size of memory M (k is fixed currently, but can vary for specific system requirement). x^{*k} is the observation. Same historical observation x might have multiple copies in memory. They are sorted so that for any $x^{*i} = x^p$ and $x^{*j} = x^q$, if $i > j$, then $p \geq q$. So in the front of the memory, we have the oldest observations, while in the end of the memory, we have most recent observations.

The model described before will be trained or updated only with the data within this memory. We design a memory update strategy, so that the model values those new observations more than those old observations, and even forget those old observations. When new observation arrives, we apply a forget (or decay) function $M = \Phi(x^*)$ to pick out old data and replace them with new observation, we can use some naive approach for the forget function (e.g. randomly picking old observations), or a probabilistic function, so the older the data, the greater chance that it would be removed: for data at index i in M, we define its probability of being forgotten as: $\psi(i) = e^{-\delta i}$. For all i, when $\psi(i) \geq \eta$, we replace them with new observation x^*:

$$\Phi(x^*) : M(i) = x^* \tag{18}$$

The decay rate δ as well as the threshold η control the speed for memory update. If δ is large or η is small, it means the old memory would be removed quick, and the model changes fast. Such memory is crucial if the target is dynamic. And for different model training methods, memory can play different roles. For point estimation method like we used for stacked autoencoder, the memory provides the data for model training and retraining. For Bayesian estimation like Gaussian process regression , the memory provides the data for calculate the likelihood, and thus influence the model update.

 ISBN: 978-1-61208-390-2

E. Surprise and anomaly detection

Attentional surprise [5] can only be defined in a relative, subjective, manner and is related to the expectations of the observer, even when derived from identical observation, same data may carry different amounts of surprise at different times. So it could be seen as the subjective measurement of information (we might call it "subjective entropy"). In this paper, surprise is defined as the KL divergence between $q(x|y)$ and $p(x|y)$. Such surprise is the trigger for model update (or retraining) as well as memory update. If surprise is larger than surprise threshold $surprise(y, x) > \xi$, it means the current model fails, and we need to update the model; also we need update the memory to include this new significant observation. In model representations, such as the stacked autoencoder used in this paper, above calculation of surprise cannot be conducted straightforwardly. Here, the surprise can be simply set as the loss value between reconstructed output and real input:

$$surprise(x) = L(x, g_{\theta'}(f_\theta(x))) \tag{19}$$

When this loss function (root mean-squared-error) output is greater than the surprise threshold ξ, it means the autoencoders cannot represent the input well at the moment. So it's necessary to update the model to fit the current input, with the data from updated memory.

This attentional surprise also enables a new method of anomaly detection. It simulates the human attention mechanism to some extent, acts as an information-processing bottleneck that allows only a small part of incoming sensory information to reach working memory and trigger model update, instead of attempting to fully process the massive sensory input. Human can maintain a certain level of alertness (e.g. when we are driving in an unknown district, we would like to pay more attention and allocation more computational resource): when it's high, more resources is prepared for attention, and even subtle signs could be detected. Similar idea is adopted here. We can find several parameters (and hyper-parameters) that provide us the chance to control surprise and model update, including: 1) Surprise threshold: model update frequency; 2) Memory size : the adaptive level of model; 3) Sensor selection parameters (pre-defines the region of sensing and change the frequency of surprise, as shown later). These top-down settings control the overall alertness of the sensing system, as well as the resource consumption. Top-down settings can be changed according to different system objectives and tasks.

F. Surprise minimization sensor selection and denoising SAE

When we only observe partial input x', the KL divergence (or surprise) is smaller than (or equal to) KL divergence with actual x. In other words, because of using partial observation, there is some space between D'_{KL} and D_{KL}. When we update the model and minimize surprise, we actually partially optimize it. Therefore, to minimize the actual KL divergence, it's necessary for sensor selection schema to keep selecting new x' to find new bound of minimization. Two different strategies can be used: one is random sensor selection, and the other is non-random sensor selection. Although random selection can reduce the space (according to our experiment shown later), it might require a large amount of iterations, especially when the sensor number is limited. So we tried to design some non-random sensor selection algorithm $s(x)_k$ that can reduce the space between D'_{KL} and D_{KL} faster. Specifically, we designed a surprise minimization sensor selection (some other methods including Markov chain are also applicable). Assume at previous observation, we have data x^1, and update model y^1 to y^2. We search for a subset of sensory space $s \subset \{m\}_k$, where:

$$surprise_{s'}(y^2, x^1) > \varrho \tag{20}$$

So it's the region that the updated model cannot well fit. We define the next sensing space s':

$$s' = (s \cup B) \cdot \omega \tag{21}$$

B is the pre-defined attentional area, which is a "spotligh" region (subset of sensory space), indicating where we are interested in ([4]). With this pre-defined region, we can locally refine the approximation, focusing computational resources to suit the task and context at hand. And ω is to generate the randomness of sensory selection. s'^* would then used as $\{m\}_k$ for $s(x)_k$.

For a stacked autoencoders, we lower the dimension of sensing by using x' as a corrupted version of x. x' is then mapped, as with the ordinary autoencoders, to a hidden representation:

$$y = f_\theta(x') = s(Wx' + b) \tag{22}$$

from which we reconstruct $z = g_{\theta'}(y) = s(W'y + b')$. z is now a deterministic function of x' rather than x. The objective function minimized by stochastic gradient descent becomes:

$$\underset{\theta,\theta'}{\operatorname{argmin}} E_{p(z,x')}[L(x', g_{\theta'}(f_\theta(x')))] \tag{23}$$

IV. RESULTS

In this section, we evaluate the methodology described in the previous section with simulated experiments. The dataset contains the simulated flow count of people for a building. We generated this data based on certain generative models (Gaussian distributions and Poisson distribution). We generate the data by random-sampling these distributions with some additional noise, and get the simulated count of people hourly for a building. We then build sensing system based on the methodology described before, which tries to understand this generative model from observation. Same challenge is posed for sensing system: the target (its generative model) can change, and the observation is not complete.

1) Change of target distribution with/without memory: In the first experiment, the parameters of generative distribution changes during the observation (e.g. the mean and variance of a Gaussian change to new values). As you can see from Figure 1, when we enable the memory window, the model can quickly capture the changed distribution and minimize the KL divergence (between estimated distribution and real distribution). Using the same model without memory,the minimization takes much longer time.

 ISBN: 978-1-61208-390-2

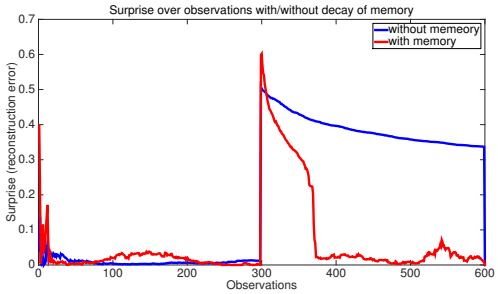


Figure 1. The target's distribution changes during the observations.

2) Change of distribution type and model representation: In the second experiment, we make more dramatic change for the generative model. So instead of changing the parameters of distribution, we change the type of distribution, from Gaussian to Poisson. As shown in Figure 2, we compared the performance of different model representations. While the Gaussian model cannot achieve KL divergence minimization. The others including dictionary learning and denoising autoencoder network work well. But we can notice that with dictionary learning method and fewer layers deep network, the KL divergence is not easy to optimize as the 5-layer deep network.

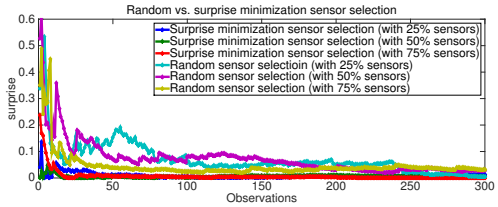


Figure 2. The distribution process changed from Gaussian to Poisson.

3) Partial observation and sensor selection: In the third experiment, we compare the random sensor selection with surprise minimization sensor selection. For a mixture Gaussian generative model with target resolution 100*100 (10000 possible sensor placements), we picked 25%, 50%, and 75% of available places for sensing, with both random and non-random sensor selection. For surprise minimization sensor selection, the initial placement is randomly picked, and then selection s' is iteratively determined by the algorithm describe before. The experiment result is shown in Figure 3, where you can find that although the random sensor selection can minimize the KL divergence between estimation and real distribution, the surprise minimization approach can achieve faster convergence.

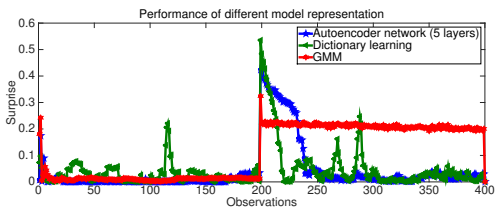


Figure 3. Surprise minimization sensor selection can better performance.

V. Conclusion

To conclude, in this paper, we 1) set a different system goal for building sensing system, which is to minimize the free energy between the model and its target; 2) adopt an attention based mechanism to detect anomaly and control the model update; 3) use a training data window as working memory mechanism; 4) utilize a deep network for model representation; 5) use partial input, based on surprise minimization sensor selection, to reduce sensing dimension.

A large amount of work is planned. we try to elaborate the components in the framework to make it more suitable for sensing intelligence: a layered or network-structured working memory will be designed to organize the data or knowledge hierarchically; model representation will be fused with external knowledges like ontology and support reasoning. Also, in the next step, we will try to apply the methodology to ECG and EEG anomaly detection. We believe for this kind of detection system, two features are highly required: firstly, it should be able to detect abnormal situation, which normally means there is something wrong or even dangerous and needs actions to be taken; secondly, it should base on personalized model, instead of an average model for large population. Therefore, the proposed methodology is believed to be suitable.

VI. Acknowledgment

This project is partly supported by the National Natural Science Foundataion of China (No. 61104215).

References

[1] Friston, K. (2010). The free-energy principle: a unified brain theory?. Nature Reviews Neuroscience, 11(2), 127-138.

[2] Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., & Manzagol, P. A. (2010). Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. The Journal of Machine Learning Research, 11, 3371-3408.

[3] Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P. A. (2008, July). Extracting and composing robust features with denoising autoencoders. In Proceedings of the 25th international conference on Machine learning (pp. 1096-1103). ACM.

[4] Whiteley, L., & Sahani, M. (2012). Attention in a Bayesian framework. Frontiers in human neuroscience, 6.

[5] Itti, L., & Baldi, P. F. (2005). Bayesian surprise attracts human attention. In Advances in neural information processing systems (pp. 547-554).

[6] Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. Science, 229(4715), 782-784.

[7] Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis?. Trends in cognitive sciences, 10(7), 301-308.

[8] Barlow, H. B. (1961). Possible principles underlying the transformations of sensory messages.

[9] Olshausen, B. A., & Field, D. J. (1996). Natural image statistics and efficient coding*. Network: computation in neural systems, 7(2), 333-339.

[10] Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. Neural computation, 7(6), 1129-1159.

[11] Hyvrinen, A., Karhunen, J., & Oja, E. (2004). Independent component analysis (Vol. 46). John Wiley & Sons.

[12] Zhong, N., Liu, J., Yao, Y., Wu, J., Lu, S., Qin, Y., & Wah, B. (2007). Web intelligence meets brain informatics. In Web Intelligence Meets Brain Informatics (pp. 1-31). Springer Berlin Heidelberg.

[13] Ma, J., Wen, J., Huang, R., & Huang, B. (2011). Cyber-individual meets brain informatics. IEEE Intelligent Systems, (5), 30-37.

[14] Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. Frontiers in human neuroscience, 4.

 ISBN: 978-1-61208-390-2

[15] Yun, Q. S., & Sun, H. (2000). Image and Video Compression for multimedia engineering. CRC.

[16] Markou, M., & Singh, S. (2003). Novelty detection: a reviewpart 2: neural network based approaches. Signal processing, 83(12), 2499-2521.

 ISBN: 978-1-61208-390-2

Towards Audio-based Distraction Estimation in the Car

Svenja Borchers, Denis Martin, Sarah Mieskes, Stefan Rieger, Cristóbal Curio, Victor Fäßler
TWT GmbH Science & Innovation
Stuttgart, Germany
email: Svenja.Borchers@twt-gmbh.de, Denis.Martin@twt-gmbh.de, Sarah.Mieskes@twt-gmbh.de, Stefan.Rieger@twt-gmbh.de, Cristobal.Curio@reutlingen-university.de, Victor.Faessler@twt-gmbh.de

***Abstract*— Distraction of the driver is one of the most frequent causes for car accidents. We aim for a computational cognitive model predicting the driver's degree of distraction during driving while performing a secondary task, such as talking with co-passengers. The secondary task might cognitively involve the driver to differing degrees depending on the topic of the conversation or the number of co-passengers. In order to detect these subtle differences in everyday driving situations, we aim to analyse in-car audio signals and combine this information with head pose and face tracking information. In the first step, we will assess driving, video and audio parameters reliably predicting cognitive distraction of the driver. These parameters will be used to train the cognitive model in estimating the degree of the driver's distraction. In the second step, we will train and test the cognitive model during conversations of the driver with co-passengers during active driving. This paper describes the work in progress of our first experiment with preliminary results concerning driving parameters corresponding to the driver's degree of distraction. In addition, the technical implementation of our experiment combining driving, video and audio data and first methodological results concerning the auditory analysis will be presented. The overall aim for the application of the cognitive distraction model is the development of a mobile user profile computing the individual distraction degree and being applicable also to other systems.**

Keywords-distraction; auditory; automotive; driver; cognitive model.

I. INTRODUCTION

Distraction during driving leads to a delay in recognition of information that is necessary to safely perform the driving task [1]. Thus, distraction is one of the most frequent causes for car accidents [2][3]. Four different forms of distraction are distinguished, although not mutually exclusive: visual, auditory, bio-mechanical (physical), and cognitive. Human attention is selective and not all sensory information is processed (consciously). When people perform two complex tasks simultaneously, such as driving and having a demanding conversation, there is an attention shift. This kind of attention shifting might also occur unconsciously. Driving performance can thus be impaired when filtered information is not encoded into working memory and thus critical warnings and safety hazards can be missed [4]. Sources for distraction of the driver can be located within and outside of the car. The continuous identification of the driver's degree of distraction could enhance safety by allowing adaptive and cooperative task automation using, e.g., advanced driver assistance systems.

Here, we will focus on in-vehicle information. This includes, but it is not limited to, in-car audio recordings and behavioural data from the driver. Multimodal data integration and synchronization is mandatory for the tool to produce meaningful results. Acoustic scene analysis comprising the detection of the number of speakers, the degree of emotional content, information about the driver's involvement in the conversation (e.g., whether the driver himself is speaking), is to be employed for the prediction of the driver's degree of distraction. In addition, eye-tracking signals, such as eye gaze direction and blink frequency, and face movement information, such as mouth movements and emotional reactions, can be exploited to increase the reliability of distraction prediction. A computational and empirical cognitive distraction model is developed for analysing the different signals, with the aim of computing a 'distraction degree' of the driver.

The effect of cognitive distraction on driving performance is empirically tested in a parallel task in order to assess the impact of auditory stimuli on distraction (cf. Figure 1). In a first experiment, we induce a continuous distraction condition and compare the driving parameters and in-car measurements with a control condition of focused, un-distracted driving. Analysing these results, we assess parameters responding reliably to cognitive distraction. These parameters are used as input for the cognitive model computing the degree of the driver's distraction. In a second experiment, we then induce a more naturalistic conversation condition leading to varying degrees of driver distraction. Our computational and empirical cognitive model is trained and tested in the course of this experiment. An acoustic analysis including the detection of the number of speakers, the degree of emotional content, information about the driver's involvement in the conversation (e.g., whether the

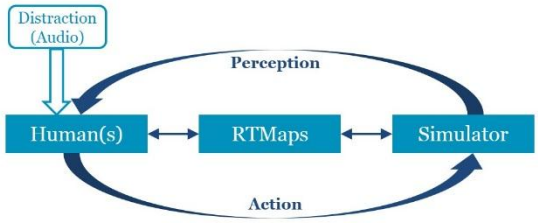


Figure 1. Perception-action loop and the influence of distraction.

 ISBN: 978-1-61208-390-2

driver himself is speaking), is used for the prediction of the driver's degree of distraction.

In Section II, the experiment is described addressing the experimental design and the features being analysed. Section III presents preliminary results of the first experiment. Section IV gives an outlook of the following analysis steps and further experiments.

II. EXPERIMENT

A. Experimental Design

In our first simulator experiment, we used a car following paradigm with the driver's task to keep the same distance to the pace car by ensuring readability of the number on the back end of the pace car. Subjects performed a practice session of three minutes driving without distraction in order to get used to the experiment and the driving simulator. The pace car drives with varying speeds between 30 and 100 km/h and brakes or accelerates 39 times during a 10 minute drive at randomly distributed locations. Some of the subjects started with the control condition, i.e., driving without distraction, of ten minutes, while other subjects started with the distraction condition. After the first condition, subjects continued with the other condition, so that each subject performed once the control and once the distraction condition. During the distraction condition, subjects were presented with simple mathematical tasks (e.g., 22+46 or 9-5) via headphones and subjects were asked to respond verbally [5][6][7]. The inter-trial interval was chosen to eight seconds. All responses were recorded.

Subjects had normal or corrected-to-normal vision and several years of driving experience each. They sat as driver in front of a large screen using a Logitech G27 game controller steering wheel with pedals (cf. Figure 2). The simulator allowed the driver to control an automatic car with the steering wheel, the gas pedal, and the brake pedal (the clutch pedal was not used). As driving simulation software, OpenDS [10] was used. Besides a custom driving task definition, minor modifications of the simulator were necessary to show brake lights of the pace car and to remotely control a software for recording videos from two web-cameras. The cameras were used to record the subject's face. One of the cameras was positioned directly in front of the subject and the other to the side front.

Figure 2. Driving simulator setup.

Synchronisation of the camera streams was guaranteed by RTMaps [9], which was remotely controlled by the OpenDS driving simulator. Facial features, mouth movements and head pose of the subject are automatically extracted to increase the reliability of distraction predictions in further analyses.

Simulator sound and the audio task were presented through a headset. Its microphone was used to record the verbal responses during the distracted condition.

B. Features

Parameters being indicators for driving performance are extracted from the driving experiments. These parameters include: distance to pace car, reaction times (both for braking and speed recovery), steering wheel jitter, and lateral position jitter. Further parameters will be evaluated for their potential use as features in the cognitive model and will be included step-by-step, e.g., head orientation (which will be relevant in conversation tasks), eye blink, and facial expressions (for emotion recognition). For conversation tasks, audio analysis will be included in the feature set of the cognitive model. In this context, features used in voice and speech recognition, such as pitch and Mel-Frequency Cepstral Coefficients (MFCC) are suitable candidates as well as derived features, such as emotional content of the utterances.

Since features used for our cognitive model will eventually come from different sources (car data, video, audio), synchronisation plays an important role. One tool allowing acquisition of multi-modal sensory data is RTMaps [9], which will be used as platform for implementing our auditory driver distraction estimation component.

III. PRELIMINARY RESULTS

Preliminary results of the driving parameters of six subjects are shown in Figure 3. All subjects showed a tendency of a larger mean distance to the pace car during distracted driving despite the explicit assignment of keeping a predefined distance determined by the readability of large numbers on the rear of the pace car. In addition, a larger variance of the distance to the pace car indicates longer reaction times for adapting to the speed of the pace car. Thus, subjects were less constantly able to keep the same distance to the pace car while they were simultaneously solving mathematical problems.

The longer reaction times are especially reflected in the deceleration case (braking instances of pace car): All subjects needed longer reaction times between the occurrence of the breaking lights of the pace car and decelerating of their own car while distracted. The variance of acceleration (including deceleration) indicates a smoother driving behavior during distracted driving (smaller spikes of the acceleration value). Together with the longer reaction times and the increased distance to the pace car, this generally shows that the subjects are more likely to drive in a safer style when cognitive workload is increased.

In conclusion, these driving parameters indicate the effectiveness of the induced distraction through the mathematical problem solving task. During our upcoming experiments, we will use these parameters to evaluate the contribution of specific auditory and facial features.

 ISBN: 978-1-61208-390-2

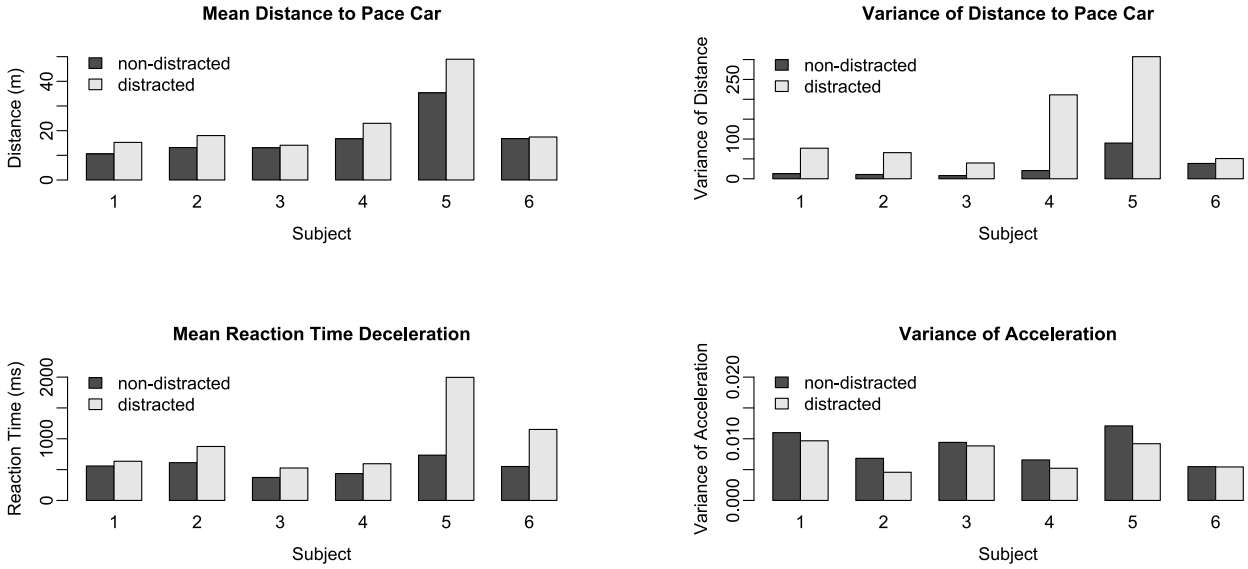


Figure 3. Driving parameters of six subjects: mean distance of the driver to the pace car, variance of distance to the pace car, mean reaction time of deceleration, and variance of acceleration.

IV. CONCLUSION AND FUTURE WORK

We have designed an experimental driving simulator setup that enables the study of behavioural and perceptual manipulations during driving, as shown by first promising quantitative results. In future work, a special focus will be on the audio scene analysis in the car interior. For this, we will extend the experimental paradigm. This will involve the driver under controlled conditions during a conversation, while monitoring her/his emotional states (i.e., through audio and facial signature analysis). First technology studies suggest that auditory features, such as pitch and MFCC are suitable candidates. An analysis of mouth movements will add information to the audio segmentation helping to identify active speakers. As platform for synchronized processing of the different data sources (audio, video, and driving performance parameters from the car), RTMaps [9] will be used (cf. Figure 4).

Several models for cue integration have been suggested for cognitive modelling of distraction. The recent dynamic Bayesian model by Liang and Lee [8] consists of a combined supervised and unsupervised learning approach. It would be interesting to extend this model with higher-level conversational cues, like the degree of estimated conversational interaction as a likely distraction measure.

Besides considering further multimodal observational cues of car passengers, especially the driver, the system should be tested and calibrated in more complex driving situations, like overtaking. Modelling current driver's task difficulty through an artificial driving model will be a further interesting research direction.

Finally, investigating strategies to support the driver by presenting and using the estimated distraction level, e.g., through visual feedback modalities or active interventions, will also be of further interest. The design of future autonomous driving systems will call for functionalities to access driver states at various automation levels.

The final technical implementation of the developed cognitive model as a mobile user profile will be further investigated throughout the project. It is likely that adaptation and life-long learning of the cognitive model will be a key feature for which a mobile application communicating with on-board car systems would be an appropriate choice.

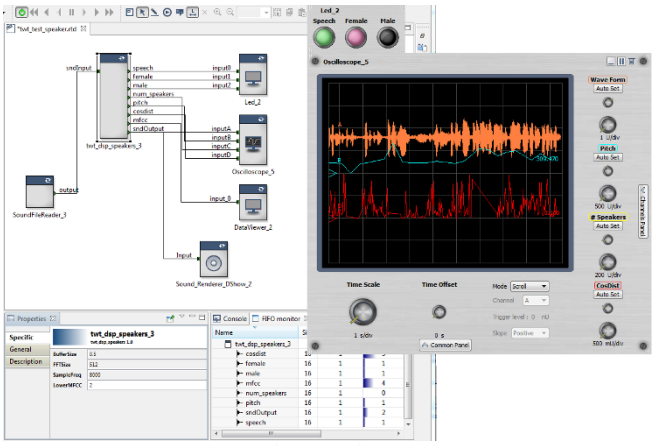


Figure 4. RTMaps for audio feature extraction.

 ISBN: 978-1-61208-390-2

ACKNOWLEDGMENT

This research has been performed with support from the EU ARTEMIS JU project HoliDes (http://www.holides.eu/) SP-8, GA No.: 332933. Any contents herein reflect only the authors' views. The ARTEMIS JU is not liable for any use that may be made of the information contained herein.

REFERENCES

[1] M.A. Regan and K. L. Young, "Driver distraction: a review of the literature and recommendations for countermeasure development," Proc. Australas. Road Safety Res. Policing Educ. Conf. 7 (v1), pp. 220–227, 2003.

[2] J. Artho, S. Schneider, and C. Boss, "Inattention and distractioin: how does the driver behave in the car?", original title: Unaufmerksamkeit und Ablenkung: Was macht der Mensch am Steuer?," Transport Research International Documentation, 2012, Online: http://trid.trb.org/view.aspx?id=1244037. Retrieved: October, 2014.

[3] T. Horberry, J. Anderson, M. A. Regan, T. J. Triggs, and J. Brown, "Driver distraction: the effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance," Accid Anal Prev 38 (1), pp. 185–191, 2006.

[4] L. M. Trick, J. T. Enns, J. Mills, and J. Vavrik, "Paying attention behind the wheel: a framework for studying the role of attention in driving," Theoretical Issues in Ergonomics Science 5 (5), pp. 385–424, 2004.

[5] M. Kutila, G. Jokela, G. Markkula, and M. Rue, "Driver distraction detection with a camera vision system," IEEE International Conference on Image Processing (ICIP 2007), vol. 6, San Antonio, Texas, USA, pp. 201-204, 2007.

[6] F. Putze, J.-P. Jarvis, and T. Schultz, "Multimodal recognition of cognitive workload for multitasking in the car," International Conference on Pattern Recognition (ICPR 2010), Istanbul, Turkey, August 2010, pp. 3748-3751.

[7] J. Harbluk, Y. Noy, P. Trbovich, and M. Eizenmann, „An on-road assessment of cognitive distraction: Impacts on drivers' visual behaviour and braking performance," Accident Analysis and Prevention, vol. 39, no.2, pp. 372-379, 2007.

[8] OpenDS, EU FP7 Project GetHomeSafe, FP7-ICT-2011-7, 288667 STREP. Online: http://opends.eu/. Retrieved: October, 2014.

[9] RTMaps, Real-Time Multimodal Applications, Intempora S.A. Online: http://intempora.com/. Retrieved: October, 2014.

[10] Y. Liang and J. D. Lee, "A hybrid Bayesian network apporach to detect driver cognitive distraction," Transportation Research Part C 38, pp. 146-155, 2014.

 ISBN: 978-1-61208-390-2

Towards Support for Verification of Adaptative Systems with Djnn

Daniel Prun, Mathieu Magnaudet, Stéphane Chatty
Université de Toulouse - ENAC
Toulouse, France
e-mail: {daniel.prun, mathieu.magnaudet, chatty}@enac.fr

***Abstract*—Djnn is a general framework dedicated to the development of complex interactive systems. We describe ongoing work aimed at developing verification mechanisms through the definition of syntax, grammar and semantics for djnn models. The results will serve to perform formal verification of interactive systems.**

Keywords-interactive system; component; control structure; model; syntax; semantic; formal verification.

I. INTRODUCTION

For more then 30 years, dedicated languages and methods have been designed and used to deal with the development of critical systems (transportation, health, nuclear and military systems). These languages and methods are used for the development of safe, functionally correct systems. For example, VHDL (VHSIC Hardware Description Language) [1] is hugely used for the development of hardware circuits, SCADE (Safety Critical Application Development Environment) [2] language is used for control and command systems.

However, highly interactive and adaptive systems have recently and progressively appeared [3], [4]. For example, air traffic control systems, surveillance systems or automotive systems have to react to many event sources: user events (from classic keyboard/mouse to more advanced interaction means such as multi-touch surfaces, gesture recognition and eye gaze), pervasive sensors, input from other subsystems, etc.

Difficulties have been observed in using existing languages and methods on these kinds of systems. Indeed, these systems require new control structures in order to manage dynamicity or to support different design styles, such as state machines and data flows, and when using existing languages this often leads to problems in the software architecture [5], [6]. We argue that part of these issues are due to the lack of a well-defined language for representing and describing interactive software design in a way that allows, on the one hand, system designers to iterate on their designs before injecting them in a development process and on the other hand, system developers to check their software against the chosen design.

This paper describes a work in progress within the development of a general framework (named Djnn) dedicated to the development of interactive systems. Section II presents the current state of Djnn and introduces requirements for its supporting systems verification. Section III discusses the early results obtained so far in the context of HoliDes (Holistic Human Factors and System Design of Adaptive Cooperative Human Machine System) project. Section IV concludes with description of future developments.

II. DJNN

Djnn [7] is a general framework aimed at describing and executing interactive systems. It is an event driven component system with:

- a unified set of underlying theoretical concepts focused on interaction,
- new architectural patterns for defining and assembling interactive components,
- support for combining interaction modalities,
- support for user centric design processes (concurrent engineering, iterative prototyping).

A. Control primitives

Djnn relies on a fundamental control primitive called "binding". A binding is a component that creates a coupling between two existing components. If there is a binding between components C1 and C2, then whenever C1 is activated, C2 is activated (C1 is called trigger and C2 is called action). A binding can be interpreted as a transfer of control, like a function call in functional programming or a callback in user interface programming. Figure 1. shows examples of binding definitions.

```
# beeping at each clock tick
binding (myclock, beep)

# controlling an animation with a mouse button
binding (mouse/left/press, animation/start)
binding (mouse/left/release, animation/stop)

# quitting the application upon a button press
binding (quitbutton/trigger, application/quit)
```

Figure 1. Examples of bindings definitions in Djnn.

Bindings can be used to derive a set of control structures required to describe interactive softwares: Finite State Machine (FSM), Connector (used to transfer data between two components), Watcher (allow to connect C1 and C2 to C3 where C3 is activated only when C1 and C2 are synchronously activated) or Switch (activates one of several components according to input data values). Figure 2. shows examples of derived control structure definitions.

 ISBN: 978-1-61208-390-2

```
# ensure that rectangle rect1 will move with
# the mouse.
connector (mouse/position/x, rect1/position/x)
connector (mouse/position/y, rect1/position/y)

# m is performed when input1 and anput2 are
# simultaneously activated
multiplication m (input1, input2, output)
watcher (input1, input2, m)
```

Figure 2. Examples of derived control structure definitions in Djnn.

FSMs are one of the most used control structures for describing user interfaces with Djnn. They contain other components named states and transitions. Transitions are bindings between two states (named origin and destination). A transition is active only when its origin is active. It behaves as a binding with a default action: changing the current state of the FSM to its destination state. Therefore, the transitions define the inputs of the state machine: the state evolves on the sequence of activation of the triggers of the transitions, and ignores events that do not match the current state. Figure 3. shows the internal behavior of a software button designed for use with a mouse: the Djnn code above implement the FSM shown at the bottom. r is the graphical representation of the button (a rectangle component).

```
component mybutton {
 rectangle r (0, 0, 100, 50)
 fsm f {
   state idle, pressed, out
   transition press(idle, r/press, pressed)
   transition trigger(pressed, r/release, idle)
   transition leave (pressed, r/leave, out)
   transition enter (out, r/enter, pressed)
 }
}
```

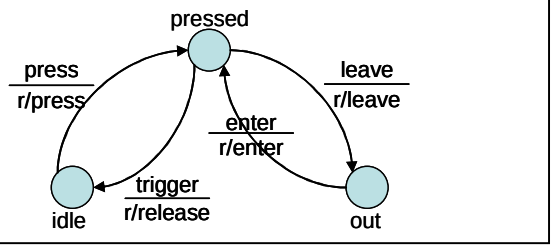


Figure 3. Example of a FSM definition.

B. An architecture of reactive components

In Djnn, every entity you can think of, abstract or physical, is a component. In addition to the control structures introduced above, Djnn comes with a collection of basic types of components dedicated to user interfaces: graphical elements, input elements (mouse, multi-touch, sensors, etc.), file elements, etc. Every type of component can be dynamically created or deleted.

To design interactive systems, components must be interconnected and organized. Interconnection is obtained with control structures, and can be performed independently of the nature and location of components. For example, a binding can connect the position of a mouse press to the position of a rectangle, so that the rectangle moves whenever the mouse is pressed. Structuration is obtained with a dedicated control structure: the parent-child interconnection that allows creating a hierarchy of components. For example, a complex graphical scene is composed of several graphical sub-components; a mouse is made of two buttons and one wheel; a FSM is made of several bindings etc. The designer can explicitly manage this tree-oriented architecture.

Combining the tree structure and the other control structures can be used for creating complex interactive behaviors and not only graphical scenes. For instance, combining FSMs by coupling their transitions, or by controlling the activation of one by a state or a transition of another, makes it possible to create complex behaviors (see example in Figure 4.). The tree structure also makes it easier to structure applications as collections of reusable components.

```
# connect “trigger” transition to a component
action “quit”
binding (mybutton/trigger, application/quit)
```

Figure 4. Example of connection with FSM.

Whenever a composite component is activated, the activation of its children components is iteratively performed. Each visited component is then activated and eventual transversal connections are activated.

C. Djnn in use: realizations and limitations

Djnn components can be created with various programming languages (Perl, Python, C, C++ or Java) or loaded from XML (Extensible Markup Languages) files. For instance, complex graphic scenes can be loaded from SVG (Scalable Vector Graphics) files. The final application is then compiled and linked with specific Djnn libraries. Dedicated available target are Windows, Linux or Mac OSX platform.

Djnn has been used for several realizations related to complex interactive systems. For example, in [8], Djnn has been used for the design and implementation of a ground control station for squads of civil Unmanned Aerial Vehicles (UAVs) (see Figure 5.). With Djnn, programming user interface adaptation comes down as a special case of programming interactive behavior. This allowed to easily implement many scenarios of adaptation, from simple state transition to complex graphical reconfigurations triggered by heterogeneous event sources. Thus, we have been able to demonstrate that Djnn provides a suitable framework to develop complex adaptive interfaces.

In [10], Djnn has been used to develop a prototype of a drawing tool overlapped on top of maps in a maritime surveillance system. The tool is used to share information between the crew during search and rescue missions. This example demonstrated how Djnn facilitates the development of user interfaces by offering a support for rapid prototyping and iterative processes.

 ISBN: 978-1-61208-390-2

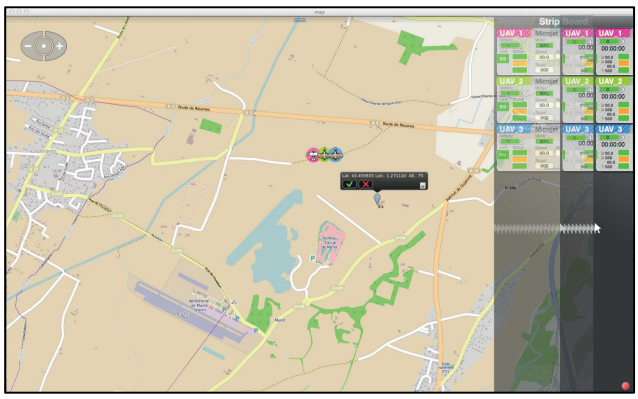


Figure 5. Djnn used for the design of a UAVs squad control.

Djnn is the visible result of an ongoing research project on programming languages for interactive systems. So far, verification of interactive systems designed with Djnn has not been studied. Focus has been put on the development of the implementation of the framework. Clearly, Djnn lacks several elements to enable the development of critical interactive systems:

- #1: a formal syntax and semantic for Djnn models,
- #2: mechanisms to translate Djnn applications into languages supporting model checking simulation or formal verification such as Event B [12] or Spin/Promela [11],
- #3: mechanisms to perform property verification directly on Djnn models.

Note that #1 is a prerequisite: without formal semantic, there is no possibility for verification. In the next section, we present our first results in this direction.

III. DJNN IN HOLIDES

The research results presented in this section are part of the HoliDes project, whose main goal is to design adaptive cooperative systems, focusing on the optimization of the distribution of workloads between humans and machines [9]. During the first year of this project, Djnn has been improved to prepare it for verification of interactive systems along two axes:

- Specification of the syntax and grammar through XML formats,
- Development of a formal semantic in Petri Nets.

A. *Syntax and grammar*

An abstract syntax and a grammar for Djnn have been defined through an XML schema. The model addresses most components available in Djnn, particularly control primitives. For example, Figure 6. contains the description of a binding and a FSM: a binding is an extension of a component containing identification of a source (“trigger”) and of a target (“action”). A FSM is an extension of a component containing a sequence of minimum of two states and a sequence of a minimum of one transition (state and transition are defined elsewhere in the XML schema).

```
<xs:complexType name="binding">
  <xs:complexContent>
    <xs:extension base="cmn:core-component">
      <xs:attribute name="source"
          type="xs:string" use="required" />
      <xs:attribute name="action"
          type="xs:string" use="required" />
    </xs:extension>
  </xs:complexContent>
</xs:complexType>

<xs:complexType name="fsm">
  <xs:complexContent>
    <xs:extension base="cmn:core-component">
      <xs:sequence>
        <xs:element name="state"
                    type="state"
                    minOccurs="2"
                    maxOccurs="unbounded" />
        <xs:element name="transition"
                    type="transition"
                    minOccurs="1"
                    maxOccurs="unbounded" />
      </xs:sequence>
    </xs:extension>
  </xs:complexContent>
</xs:complexType>
```

Figure 6. Djnn binding and FSM control structures described by the XML schema.

The main advantages provided by these definitions are:

- Definition of a well-defined model for Djnn: illicit constructs using the language can easily and automatically be detected during edition of the model thanks to the XML schema.
- Improvement of interoperability: this evolution is a first step towards the definition of a better integrated tool chain with the capability to dump a concrete GUI (Graphical User Interface) in an XML file, and conversely, to load and to execute a GUI from an XML based description.

B. *Towards a formal semantic*

Semantic of Djnn model is expressed through colored Petri Nets [13] extended with reset arcs [15]. We chose this formalism because, at a first glance, it offers good characteristics to represent both static and dynamic concerns through a state-transition semantic. It also allows to model simple data. All Djnn components are currently being individually modeled with Petri Nets. Figure 7. and Figure 8. give overviews of the semantic. The left part of Figure 7. represents a binding between a source and an action with a simple and unique transition. As a binding is a component, its interface also offers run and stop operations. The right part represents a connector: when activated, input data <X> is copied to the output. Figure 8. shows Petri Nets model of the button as defined in Figure 3.

 ISBN: 978-1-61208-390-2

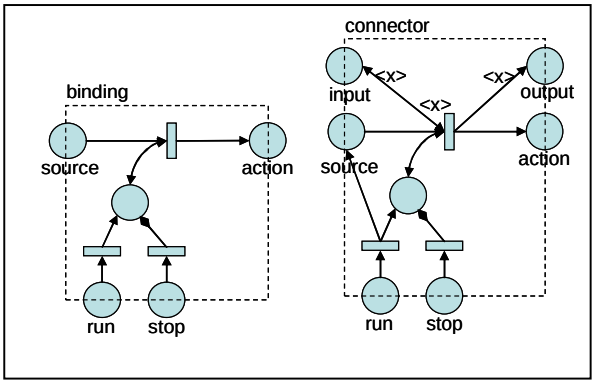


Figure 7. Models of a binding and a connector.

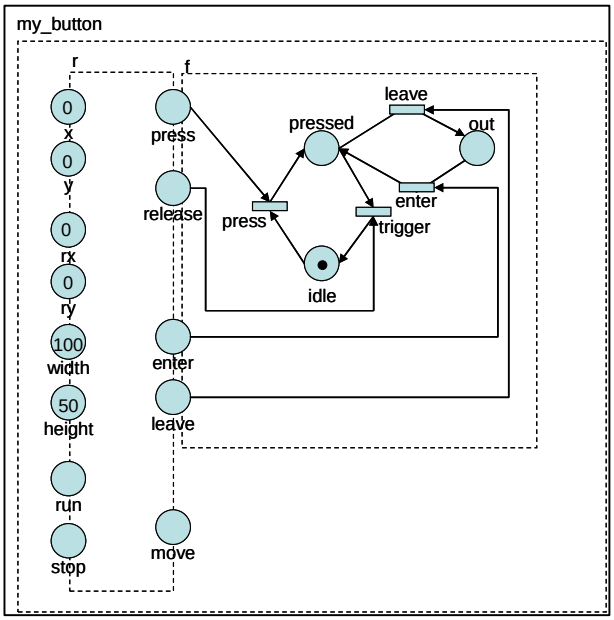


Figure 8. Model of FSM (as defined in Figure 3.

Composition of components is achieved through the merging of places of Petri Nets. This model of composition, even if it is asynchronous, seems to perform best for our purpose.

Such a formal definition of the semantic is central for verification purpose because:

- semantic of Djnn is no longer subject to misunderstandings or interpretations. A Djnn model has the same meaning for every actor in interaction with it (designer, code developer, final user, etc.);
- as the Petri Nets semantic is formal, several mathematical verifications become enabled: for example, LTL (Linear Temporal Logic) or CTL (Computation Tree Logic) properties [14], liveness or boundness properties. Moreover, translations to other languages specialized on formal verification become possible.

IV. CONCLUSION AND FUTURE PLANS

In this paper, current research related to a framework for the development and the verification of interactive safety critical systems has been presented. Although bases have already been developed (syntax and grammar through a XML schema, part of the semantic with Petri Nets), investigations remain to be done:

- So far, Petri Nets have showed their capability to model Djnn elements and mechanisms but some further analysis must be done on dynamic aspects of Djnn (creation/destruction of components).
- Use of the Petri Nets models to perform verification through simulation or through model analysis.
- Connections with tools specialized in formal verification.

Application on some real use cases, hopefully brought by HoliDes project, are also planned for the next phases.

V. ACKNOWLEDGMENTS

This research has been performed with support from the EU ARTEMIS JU project HoliDes (http://www.holides.eu/) SP-8, GA No.: 332933. Any contents herein reflect only the authors' views. The ARTEMIS JU is not liable for any use that may be made of the information contained herein.

REFERENCES

[1] "VHDL Language Reference Manual", IEEE Std 1076-2008.

[2] Scade homepage, http://www.esterel-technologies.com, [retrieved: 02, 2015]

[3] L. Bass et al. "The Arch model: Seeheim revisited", CHI'91 User Interface Developers Workshop, Apr. 1991.

[4] G.E. Pfaff, "User Interface Management Systems," Eurographics Seminars, Springer-Verlag, 1985.

[5] B. A. Myers, "Separating application code from toolkits: Eliminating the spaghetti of callbacks," In Proc. UIST, 1991, pp. 211–220, Addison-Wesley.

[6] B. A. Myers and M. B. Rosson, "Survey on user interface programming," In Proc. CHI, 1992, pp. 195–202, ACM Press.

[7] Djnn project homepage, http://djnn.net, [retrieved: 02, 2015].

[8] M. Magnaudet and S. Chatty, "What should adaptivity mean to interactive software programmers?" EICS 2014, ACM SIGCHI, Rome, Italy, Jun 2014, pp 13-22.

[9] HoliDes (Holistic Human Factors and System Design of Adaptive Cooperative Human Machine System) R&D project www.holides.eu, [retrieved: 02, 2015].

[10] C. Letondal, P. Pillain, E. Verdurand, D. Prun, and O. Grisvard, "Of Models, Rationales and Prototypes: Studying Designer Needs in an Airborne Maritime Surveillance Drawing Tool to Support Audio Communication," In Proc. of BCS HCI, ACM, 2014, pp. 92-102.

[11] G.J. Holzmann, "The Spin Model Checker: Primer and Reference Manual," 2003, Addison-Wesley.

[12] J.-R. Abrial, "Modeling in Event-B: System and Software Engineering," May 2010, ISBN: 9780521895569.

[13] K. Jensen, "Coloured Petri Nets," Berlin, Heidelberg, 1996, ISBN 3-540-60943-1.

[14] C. Baier and J.-P. Katoen, "Principles of Model Checking," 2008, The MIT Press.

[15] C. Dufourd, A. Finkel, and P. Schnoebelen, "Reset nets between decidability and undecidability," In 25th ICALP, vol. 1443 of LNCS, Springer, July 1998, pp. 103-115.

 ISBN: 978-1-61208-390-2

Adaptive Human-Automation Cooperation: A General Architecture for the Cockpit and its Application in the A-PiMod Project

Denis Javaux*, Florian Fortmann†, Christoph Möhlenbrink‡
*Symbio Concepts & Products, Bassenge, Belgium, Email: denis.javaux@symbio.pro
†OFFIS, Oldenburg, Germany, Email: florian.fortmann@offis.de
‡DLR, Braunschweig, Germany, Email: christoph.moehlenbrink@dlr.de

***Abstract*—The design of future aircraft cockpits will be based on a cooperative team perspective of the human crew and the automation. The team perspective requires to rethink the interaction between the human crew and the automation. It further requires to develop a human-machine system architecture that supports this perspective. This paper describes a general architecture for adaptive human-automation cooperation in the cockpit. It relies on an analysis of the nature of the flight for better integration of the crew and cockpit automation in the joint, cooperative and adaptive completion of the flight. The architecture has been instantiated in the European project A-PiMod, which aims at developing adaptive multi-modal cockpits to support the interaction between the human crew and the automation.**

Keywords–Human-machine cooperation; adaptive systems; automation design.

I. INTRODUCTION

Improving aircraft safety is one of the main challenges to cope with the expected increase of future air traffic. Modern aircraft are highly complex socio-technical systems, comprised of a human crew and the automation. The trend towards more automation has changed the task of the human crew from manual to supervisory control [1], which has led to novel error sources. Studies have shown that 60-80% of aviation accidents are caused by human errors [2] [3], such as automation surprises [4], opacity [5], erroneous mental models [6], degraded situation awareness [7], and out of the loop problems [8]. These problems have significantly contributed to major incidents and accidents, despite the high level of training of the human crew. This hints at a disconnection between the human crew and the automation, inherent to the way automation is currently designed. To address this design problem, several authors [9] [10] argue that automation should be seen as a team player. Both, the human crew and the automation should constitute a deeply integrated team that achieves the mission safely and adaptively in all circumstances.

In an unpublished work, the first author has been investigating for years possible new architectures for human-automation cooperation in the cockpit. Results showed that cooperative human-automation systems are implicit in many operational settings, including the cockpit, and that many Human Factors issues plaguing them were likely the results of the non-acknowledgment of that implicit nature. These systems should be understood, modeled and explicitly, purposefully and rationally designed as explicit human-automation cooperative systems.

In this paper, we propose a general architecture for adaptive human-automation cooperation that precisely shows how to build such a human-automation team, based on an in-depth analysis of (1) the nature of the flight and its execution, and (2) adaptive, cooperative human-machine systems. The architecture is currently instantiated under the umbrella of the European Research & Development project A-PiMod [11], which consists of a consortium of partners from the research and the industry.

This paper is structured as follows. In Section II, we elaborate on the nature of the flight and its execution. In Section III, we describe the general architecture. In Section IV, we describe the application of the architecture within the A-PiMod project. In Section V, we finally provide a conclusion on our work.

II. THE NATURE OF THE FLIGHT AN ITS EXECUTION

In order to analyze, model and design an airliner cockpit as an explicit human-automation cooperative system it is necessary to understand the very nature of the flight, in abstract and purely functional terms. What is a flight and what does it entail from the point of view of a controller in charge of executing that flight.

A. Mission Level Tasks: What the Flight is About

Performing a flight is about going from origin to destination, safely and efficiently, while adapting to current contingencies (weather changes, system failures, crew incapacity, etc.). Indeed, the flight plan (F-PLN) cannot always be flown as intended and has to be altered, radically modified, or even aborted. Therefore, the F-PLN is a dynamic structure that is permanently adapted. It is akin to the most general flight task, that has to be performed by the aircraft (A/C). We call this task the Mission Level (ML) task. The ML task can be decomposed into several ML subtasks, that correspond to the different flight phases. Figure 1 shows the general structure of the ML task, represented as a graph.

Performing the ML task consists of progressing over that graph and taking appropriate alternative branches if necessary (e.g., to return to departing airport). The graph (the mission) is executed by some form of control logic. This control logic may be on the ground (e.g., if the A/C is a remotely controlled unmanned aerial vehicle), on-board (e.g., if the aircraft is completely autonomous, manually controlled by one to many human pilots, under shared control of a team of some form of automation and a human crew), or distributed between ground and on-board entities (e.g., if the A/C is a remotely controlled unmanned aerial vehicle with some form of automation on-board). All these control logic designs are functionally equivalent and perform the same functions. They differ solely in

 ISBN: 978-1-61208-390-2

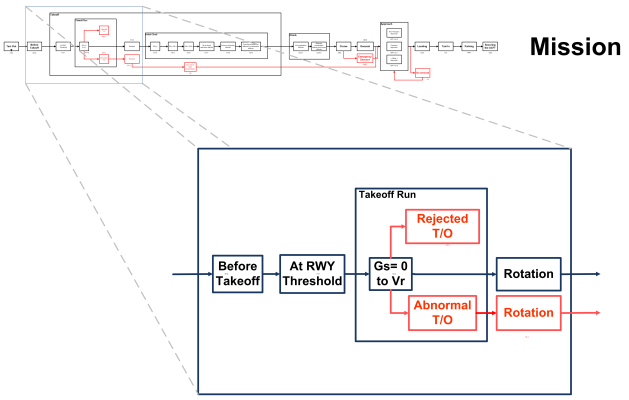


Figure 1. Representation of the ML task as a graph. The graph shows the structure of the flight, including its phases, and adaptation to contingencies (highlighted in red).

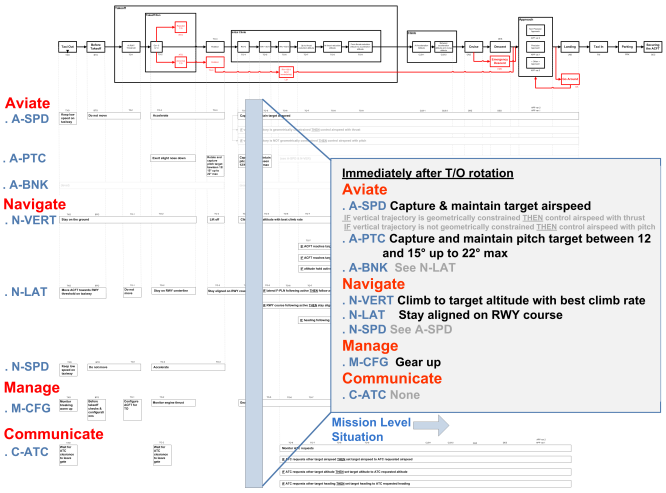


Figure 2. Translation of current ML subtask into mission execution tasks.

terms of their implementation (human or automation) and their physical distribution (ground or on-board).

The general architecture proposed in this paper relies on the deep recognition of that functional equivalence. It prompts at understanding contemporary cockpits as dynamic functional systems that implement the control logic needed to execute the mission, including its monitoring and adaptation whenever necessary.

B. Cockpit Level Tasks: What the Cockpit has to do

In modern commercial aviation, the execution, monitoring and adaptation of the ML task is achieved by the human crew and automation, seen as a set of cooperative human and machine agents in the cockpit. To achieve the ML task, the cockpit has to perform the following three types of Cockpit Level (CL) tasks:

(1) Mission (F-PLN) Monitoring and Adaptation Tasks: The F-PLN is not a static structure. During a given flight, it is frequently adapted, to cope with Air Traffic Control (ATC) requests, or because of contingencies that induce its modification (e.g., bad weather at destination airport, and weather avoidance during the flight). The cockpit therefore also has to permanently monitor the mission, assess the current circumstances, and decide if changes need to be made to the F-PLN. These are mission monitoring and adaptation tasks.

(2) Mission (F-PLN) Execution Tasks: The cockpit has to take as input the current ML subtask and perform corresponding lower level mission execution tasks (flight control, A/C configuration, interaction with ATC, etc.). The mission execution tasks can be organized along the familiar categories ”Aviate”, ”Navigate”, ”Communicate” and ”Manage”. In the schema of Figure 2, the ”Aviate” category contains, e.g., three flight control tasks: airspeed control (A-SPD), pitch control (A-PTC), and bank control (A-BNK). Similarly, the ”Navigate” category incorporates the flight navigation tasks vertical navigation (N-VER), lateral navigation (N-LAT), and ground speed control (N-SPD). The graph exactly shows which subtasks have to be executed to complete the mission. For example, at rotation during takeoff, A-SPD requests an accelerating airspeed, A-PTC a rotation to reach a target pitch between 15° and 22°, and A-BNK a bank angle of about 0° (wings level). Thus, taking the mission (F-PLN) as input, specific mission execution tasks are derived. These tasks specify what the A/C has to do to execute the current mission. If ever the mission needs to be modified, the mission execution tasks are dynamically updated to reflect the new mission.

(3) Task Distribution Tasks: The CL tasks have to be performed by the given control logic, in this case the aircraft cockpit as a whole, including the human crew and the automation. Some processing is needed to decide who will perform them. In today’s cockpits for example, mission (F-PLN) execution is mostly (99% of the time) performed by the Auto Flight System (AFS). Mission monitoring and adaptation is mostly achieved by the human crew, with some assistance from the Flight Management System (FMS). This allocation, however, is very static and this is detrimental to the cockpit adaptability and resilience. One of the objectives of the general architecture proposed in this paper is to permanently suggest a suitable distribution of CL tasks, based on the state, capabilities and workload of the agents, especially the human crew. This distribution is achieved by the task distribution tasks.

C. Agent Level Tasks: Executing Cockpit Level Tasks

CL tasks are executed by the agents in the cockpit after being distributed to them. We call the concrete distribution of these tasks the Agent Level (AL) tasks. As mentioned above, F-PLN execution in contemporary cockpits is mostly achieved by the automation, F-PLN monitoring and adaptation by the human crew with some assistance by the automation, and task distribution is achieved by the human crew and by mode control logics inherent to the automation. The AFS, including the autopilot (AP), Auto-Throttle (ATHR) and FMS indeed have specific internal modes. Each mode achieves one or more CL tasks (e.g., the speed mode on Airbus aircraft regulates airspeed). These systems have some mode transition logics that define how they switch between modes, and therefore between tasks. They implement a form of implicit and automatic task (re-) distribution in the cockpit.

One of the objectives of this paper is to suggest that there is a disconnection between these implicit task distributions by automation and the ones to which the human crew contribute,

 ISBN: 978-1-61208-390-2

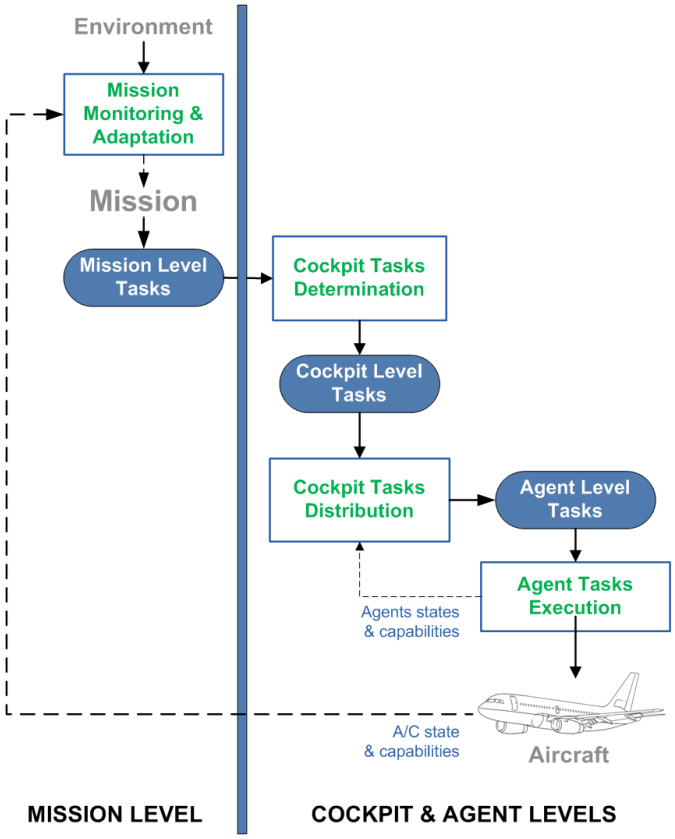


Figure 3. General Architecture for Adaptive Human-Automation Cooperation.

something typically manifested by the now infamous automation surprises. Automation surprises occur, e.g., because the human crew and the automation do not cooperatively define how tasks should be distributed, something the proposed architecture aims to improve upon.

III. An Architecture for Mission, Cockpit and Agent Level Tasks

From this deeper understanding of the F-PLN and its execution, a general architecture for the distributed, cooperative and adaptive execution of ML, CL, and AL tasks in modern aircraft cockpits can be derived. It allows the execution by a cooperative human-automation system. The architecture goes beyond the scope of cockpits and is usable in most frameworks where a mission has to be achieved by a collective of human and machine agents. The architecture derives very naturally from the distinction above between ML, CL, and AL tasks (see Figure 3). It relies on four components that process them specifically:

(1) Mission Monitoring and Adaptation: The progress over the Mission (F-PLN) is assessed and as well as the context in which the Mission is executed. Context includes, e.g., weather near the A/C, along the foreseen F-PLN and at destination, the state of installations at destination (e.g., landing assistance systems such as the ILS), and the availability of the intended runway. It also includes the state of the A/C itself. Any abnormal contingency in the context will prompt for an adaptation of the Mission (F-PLN) (e.g., to bypass bad weather) or more drastic changes such as diversion to another airport. The output of this component is the current Mission task, a F-PLN that is always safe and flyable, based on the initial FPLN and permanent adaptation to external contingencies (and ATC requests).

(2) Cockpit Tasks Determination: The tasks the cockpit seen as a whole has to perform at all times are produced. There are two types of cockpit tasks: 1) mission independent tasks: these tasks have to be performed permanently by the cockpit: Mission (F-PLN) monitoring and adaptation tasks, Task distribution tasks and Agent level tasks: Executing cockpit level tasks; 2) mission dependent tasks: they are directly derived from the current Mission Task (a safe and flyable F-PLN) and implement the flight plan in terms of concrete actions. A possible strategy for producing them is to rely on the dimensions suggested in the Figure 2 (Aviate, Navigate, Communicate, Manage), which allow easily translating the F-PLN into lower level tasks such as having specific airspeed, ground speed, thrust, heading, pitch, bank, altitude, descent angles, communication occurrences, and actions on A/C systems such as the gear, flaps, and slats.

(3) Cockpit Tasks Distribution: Once cockpit tasks have been determined, they have to be distributed to the different agents available in the cockpit, that is the crew (PF and PNF) and various automated systems (e.g., AFS, automated fuel monitoring, etc.). This therefore takes as input the cockpit tasks to distribute and the state and capabilities of the agents in the cockpit (e.g., vigilance state of the crew, state of fatigue, situation awareness, workload, state of automation, etc.). In a normal flight today, most of the time, the Mission (F-PLN) execution tasks are assigned to the AFS, while most mission independent tasks are assigned to the crew with some assistance of cockpit automation systems (envelope protections, collision avoidance systems, etc.). This is particularly true of the Mission (F-PLN) monitoring and adaptation tasks.

(4) Agent Tasks Execution: The tasks assigned to the agents are then executed by them. Thus Mission (F-PLN) execution is today mostly achieved by the AFS and the other mission independent tasks by the crew with assistance from some automated systems.

Each of the four components in the architecture should be seen as a cooperative human-machine system in its own respect: each of them is made of human and machine agents cooperating to perform the work of the component. This is one of the key ideas of the architecture. Another important idea is that the agents in question are functional agents. A single physical agent can embody several functional agents (e.g., participate to several functional agents, in more than one, and possibly all of the four components). In the extreme case of fully manual flight, there are only four functional agents, one for each component, and they are all achieved by a single physical agent: the pilot. The pilot superposes the four functions. As will be seen later, in modern cockpits, the crew also superposes participation to all four components, into two single individuals: the pilot flying (PF) and pilot non-flying (PNF). This superposition idea is the key to cooperative system design.

As seen in Figure 3, the whole architecture is mostly divided into two sections: the Mission Level is about everything

 ISBN: 978-1-61208-390-2

occurring ”outside the cockpit” and is about determining the Mission the logic in control of its execution has to achieve (cf. the notion of functional equivalence above. Other architectures, e.g., an autonomous drone, would be absolutely identical here). The Cockpit and Agent Levels are about the peculiar proposed implementation and how it achieves the mission (in term of the functional approach, all implementations would differ here).

Two main control loops exists within the architecture (beside the control loop closed by the agents on the A/C): a loop to monitor and adapt the mission and a loop to monitor and adapt the task distribution within the cockpit. They provide the adaptive capabilities expected from the architecture: to external circumstances (e.g., change of F-PLN in case of weather change) and to internal circumstances (e.g., change of task distribution because of high workload for one of the human crew). The whole architecture hints are the inherently mission-oriented character of future cockpit architectures. Future cockpits should be about the safe and adaptive completion of the mission, by an adaptive and cooperative system of human and machine agents (on-board, and/or on the ground).

IV. Application of the Architecture within the A-PiMod Project

A-PiMod (Applying Pilot Models for Safer Aircraft) is a European project that aims at developing adaptive automation for a multi-modal cockpit. Indeed, todays automation is indifferent to the emotional and cognitive state of the crew. Automation only supports the crew based on explicit and static task assignments, with no adaptive capabilities, even though it is capable of higher or lower levels of support if needed or when the capabilities of the crew are challenged. A novel approach to adaptive automation is needed and it must be applicable for real-time operations. Automation should be seen as a partner in the global endeavor of flying the aircraft with the human crew; they should adapt to each other and to the context, aiming at maintaining safety at all times as a team. The objective of A-PiMod is therefore to provide adaptive task distributions between the crew and automation, based the crew’s state (workload, situation awareness, or fatigue, etc.) and behavior (e.g., a critical task that is not performed by the crew will be taken over by automation). In order to provide means of improvement to the safety of flight, especially in times of continuously increasing performance levels, automation and information provision to the flight deck, A-PiMod works on adaptive automation and an adaptive multi-modal cockpit.

The A-PiMod consortium comprises eight European partners from 6 countries. A-PiMod will last from September 2013 to August 2016. A-PiMod is followed yearly by an advisory group of highly qualified pilots, Human Factors researchers and industry experts. During the project, the partners have defined a framework for adaptive automation based on the general architecture for adaptive human-automation cooperation shown in Figure 3. The A-PiMod project includes partners with competences for the design of a multi-modal cockpit, for building crew state inference models, with expertise in real-time risk assessment and training. The A-PiMod project is continuously going along with validation activities and addressing safety as an operational concept.

The A-PiMod architecture is based on 8 components and 2 separate software modules, arranged within the three macro-

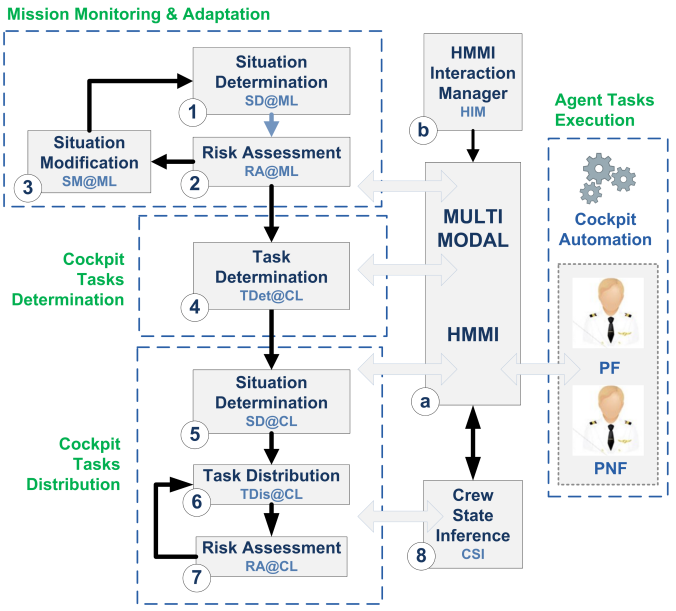


Figure 4. A-PiMod architecture.

components (Mission Monitoring and Adaptation, Cockpit Tasks Determination, Cockpit Tasks Distribution, and Agent Tasks Execution) of the general architecture previously shown in Figure 4. A peculiarity of the A-PiMod architecture, inherited for the general architecture, is the inherently cooperative nature of the components: the components are not only software. In the A-PiMod architecture, a component is made of a software module and of the human crew (PF and PNF). Thus, each component is a small cooperative system in itself. These components are accompanied by two exclusive ”software only” modules, which realize the interaction between the human crew and the automation. In the following, the components (1-8) and the exclusive software modules (a and b) of the instantiated A-PiMod architecture are described.

(1) Situation Determination at Mission Level: The situation determination at Mission Level component (SD@ML) is in charge of determining the current state of the mission and the context in which it is executed. This includes the progress on the F-PLN (mission phase/sub-phase), the state of the A/C and its systems and the environment in which the A/C operates (e.g., weather, runway availability at destination airport, and ATC).

(2) Risk Assessment at Mission Level: The risk assessment at Mission Level component (RA@ML) is in charge of determining the risk of not being able to achieve the mission as intended (e.g., can the current F-PLN be flown safely to the destination?). It uses as input the output of the situation determination at Mission Level component (SD@ML).

(3) Situation Modification at Mission Level: If the risk level is deemed unacceptable, control is passed to the situation modification at Mission Level component (SM@ML). This component is in charge of reducing the risk associated with the current situation to an acceptable level. For example, this will entail solving any threatening issue with the A/C systems (e.g., engine fire) or modifying the F-PLN, e.g., to avoid bad weather, or chose an alternate destination.

 ISBN: 978-1-61208-390-2

(4) Task Determination at Cockpit Level: When the risk level is back to acceptable levels, the cockpit level tasks for the current situation are determined by the task determination at Cockpit Level component (TDet@CL). This includes all F-PLN execution tasks, all F-PLN monitoring and adaptation tasks and all task distribution tasks (e.g., choosing which cockpit agent has to do what).

(5) Situation Determination at the Cockpit Level: The situation determination at the Cockpit Level component (SD@CL) then assesses the state of the cockpit, in particular the state of the agents, human and machine (e.g., crew and automation), in terms of availability and current capabilities (e.g., crew is fatigued).

(6) Task Distribution at Cockpit Level: The cockpit level tasks and the cockpit state are then processed by the task distribution at cockpit level component (TDis@CL) to produces one or more tentative distribution of the tasks to the cockpit agents.

(7) Risk Assessment at the Cockpit Level: The risk assessment at the Cockpit Level component (RA@CL) then assesses the risk(s) associated with these distribution(s). This risk evaluation is based on the state, workload and capabilities of the agents (e.g., the crew and automation, such as the AFS). A task distribution will only be selected if the associated risk is deemed acceptable, and if several distributions exists, the one with the lowest risk will be selected. If no acceptable distribution can be found, this information is passed back to the task determination at Cockpit Level component (TDet@CL) to state that the requested set of cockpit level tasks cannot be achieved safely by the cockpit. This information is then transferred to the situation modification at Mission Level component (SM@ML), which will typically produces an alternate F-PLN (because the previous one could not be flown safely by the cockpit). For example, this would happen if one of the crew was incapacitated and the cockpit workload was likely to be so high (e.g., due to bad weather at destination airport) that a diversion to another airport would be safer (and flyable by the diminished cockpit).

(8) Crew State Inference: Adaptivity of the task distribution within the cockpit heavily depends on the capability of the task distribution and risk assessment components at the Cockpit Level (TDis@CL and RA@CL) to know the current and future state of the crew. This information is provided by an eighth component, the Crew State Inference component (CSI). The CSI permanently monitors the crew and infers their current state, such as vigilance, workload and situation awareness. As all components in the architecture, the component is made of the crew and of a sophisticated software module. The software module produces its own inferences but the crew can alter them if needed, or even indicate they are fatigued or incapacitated before the module detects it. The CSI module infers intentions, situation awareness, and workload, using a combination of well-studied technologies, notably Bayesian and cognitive models.

(a) Human-Machine Multi-Modal Interface: The Human-Machine Multi-Modal Interface (HMMI) plays a role in communicating information to the crew and for supporting the interaction and cooperation between the crew and the software modules within the components (e.g., supporting the joint modification of the F-PLN by the SM@ML module and the

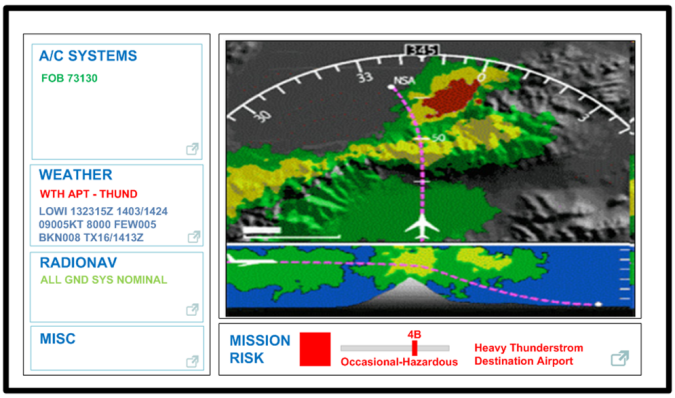


Figure 5. Paper prototype of the Mission Level display. It shows a map with the current flight plan and weather information, and associated mission risks.

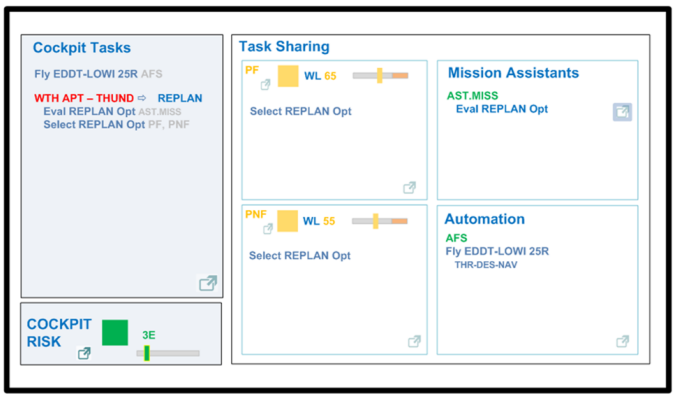


Figure 6. Paper prototype of the Cockpit Level display. It shows the current task distribution between the human crew and the automation, the workload for PF and PM and the associated cockpit risks.

crew). Besides the traditional input modalities, the HMMI relies on speech, gesture, and touch input. Further eye movements are recorded as a measure of attention. In A-PiMod, the HMMI for presenting information and supporting human-automation cooperation is currently under development. The current tentative solutions shown in Figure 5 and Figure 6 rely on two dedicated displays: the ML display and the CL display. The two displays support the interaction and cooperation between the human crew and all software modules in the architecture. They support the whole collaborative execution, monitoring and adaptation chain of the mission by the human crew and the modules, providing adaptive task distribution at all stages, based on the state, workload and capabilities of the cockpit agents. This provides adaptive capabilities to the HMMI, where the goal is to ensure the information displayed is indeed processed as intended.

(b) HMMI Interaction Manager: To drive the HMMI finally, the A-PiMod architecture sports the HMMI Interaction Manager. The HMMI Interaction Manager is a pure software module (e.g., not a component in which the crew intervene). Its task is to handle the interaction between the modules and the human crew, in particular by considering the current state and actions of the crew (provided by the CSI). The HMMI Interaction Manager gets requests from the modules to display information (on the ML and CL displays). It then displays the

 ISBN: 978-1-61208-390-2

corresponding information, but does not stop there. It starts monitoring if the information is perceived by the human crew (or if a task to be triggered by that information is executed). If not, it will enter an escalation strategy, e.g., by enhancing the salience of the information, and possibly resorting to alarms if necessary.

To implement the A-PiMod architecture, the project partners have specified and developed a series of software modules that act as team players with the crew to perform and manage the flight. These modules are integrated into several demonstrators that are used to conduct extensive validation sessions at DLR premises, of the underlying adaptive automation concepts and of the demonstrators themselves. Many of the modules in the A-PiMod architecture are driven by rule production systems. This allows for real-time behavior, deterministic execution and their certifiability.

V. CONCLUSIONS

We believe the general cockpit architecture above and its instantiation in the framework of A-PiMod has the potential to improve the safety of future aircraft. It provides a complete adaptive execution and adaptation chain for the mission, based on 4 main core components (instantiated into 8+2 components and software modules in A-PiMod). The architecture smoothly adapts to changes at the Mission level (e.g., bad weather) and at the Cockpit and Agent levels (e.g., incapacitation). In A-PiMod the later is achieved through a multi-modal HMMI and a CSI module. Each component is a fully cooperative system made of the crew and dedicated software agents (modules in A-PiMod). This allows task sharing within the components that range from full manual execution to full automatic execution. The software agents (modules) and the crew basically contribute the same tasks. The software agents can be seen - and should be designed - as cognitive agents [12]. This makes human interaction and cooperation with them far easier and robust. The architecture integrates all state transitions for automation systems in the cockpit into a single task distribution component, itself dealing with the allocation of tasks to the crew and automation. This makes the global behavior of automation far more integrated, easier to develop and debug, and synchronized with the crew during operators (reduction of automation surprises). The architecture provides a framework for aircraft that can be flown in full (or assisted) manual or full automatic control, with many intermediary configuration between which it is easy and safe to transition in flight. The architecture is also ideal for progressively, non disruptively and safely developing aircraft that implement more automation. Given the cooperative nature of the components in the two architectures (general and A-PiMod) it is possible at any time to revert in-flight to mixed or full manual modes.

ACKNOWLEDGMENT

The A-PiMod project is funded by the European Commission Seventh Framework Programme (FP7/2007-2013) under contract number: 605141 Project A-PiMod.

REFERENCES

[1] T. Sheridan, Telerobotics, Automation, and Human Supervisory Control. MIT press, 1992.

[2] C. Billings, Flight Deck Automation: Promises and Realities. NASA Ames Research Center, 1989, ch. Toward Human Centered Automation, pp. 167–190.

[3] R. Amalberti, “Automation in aviation: A human factors perspective,” Handbook of aviation human factors, 1999, pp. 173–192.

[4] N. B. Sarter, D. D. Woods, and C. E. Billings, “Automation surprises,” Handbook of human factors and ergonomics, vol. 2, 1997, pp. 1926–1943.

[5] C. Billings, Aviation automation: the search for a human-centered approach, ser. Human factors in transportation. Lawrence Erlbaum Associates Publishers, 1996.

[6] N. B. Sarter and D. D. Woods, “How in the world did we ever get into that mode? mode error and awareness in supervisory control,” Human Factors: The Journal of the Human Factors and Ergonomics Society, vol. 37, no. 1, 1995, pp. 5–19.

[7] M. Endsley, “Automation and situation awareness. automation and human performance: Theory and applications,” Parasuraman, R., Mouloua, M.(eds.), 1996, pp. 163–181.

[8] M. R. Endsley and E. O. Kiris, “The out-of-the-loop performance problem and level of control in automation,” Human Factors: The Journal of the Human Factors and Ergonomics Society, vol. 37, no. 2, 1995, pp. 381–394.

[9] F. Flemisch, S. Meier, J. Neuhöfer, M. Baltzer, E. Altendorf, and E. Özyurt, “Kognitive und kooperative systeme in der fahrzeugführung: Selektiver rückblick über die letzten dekaden und spekulation über die zukunft (cognitive and cooperative systems in vehicle control: Review of past and speculation of the future),” Kognitive Systeme, 2013.

[10] R. Parasuraman, M. Barnes, K. Cosenzo, and S. Mulgund, “Adaptive automation for human-robot teaming in future command and control systems,” DTIC Document, Tech. Rep., 2007.

[11] A-PiMod, “Applying pilot models for safer aircraft,” February 2015. [Online]. Available: http://www.apimod.eu

[12] R. Onken, “Cognitive cooperation for the sake of the human-machine team effectiveness,” DTIC Document, Tech. Rep., 2003.

 ISBN: 978-1-61208-390-2

Multidimensional Pilot Crew State Inference for Improved Pilot Crew-Automation Partnership

Stefan Suck and Florian Fortmann
OFFIS - Institute for Information Technology
Oldenburg, Germany
Email: stefan.suck, florian.fortmann@offis.de

***Abstract*—Automation is a substantial technology of modern aircraft. Even though automation has significantly improved aviation safety, insufficient partnership between the pilot crew and the automation, and confusion over the status of the automation is still a problem. The European project A-PiMod addresses these problems by developing a virtual crew member, which takes the position of classical aircraft automation. As part of the crew, the virtual crew member must be able to anticipate the internal states of the human crew members. This ability helps, e.g., to improve the task share in the cockpit by means of dynamic adaptions of task distributions. In this paper, we present the concept of the A-PiMod pilot model, which will be used for inferring the internal state of the human crew members. The internal state is composed of different sub-states, which have been defined during the initial phase of the project. The addressed sub-states are situation awareness, workload, and intentions. The target states will be inferred based on real-time data about the mission, tasks, and pilot behaviors, including what they say, where they look at, and how they act.**

Keywords–Human-Machine Cooperation; Cognitive Model; Aircraft Crew.

I. INTRODUCTION

Automation is a substantial technology of modern aircraft [1]. Automation accomplishes (partially or fully) a task that was previously carried out (partially or fully) by a human operator [2]. Overall, automation has significantly improved aviation safety [3]. However, after many years of automation it turned out that this technology is like a two-edged sword. It has been shown that there are several pitfalls associated with automation [4], such as insufficient partnership between the pilot crew and the automation, and confusion over the status of the automation. These pitfalls refer, e.g., to the lack of communicating internal states, including the situational picture and the intents of the pilot crew and the automation. Insufficient partnership between the pilot crew and the automation has led to several accidents in the past. A well-studied example is the crash of China Airlines Flight 140, which can be attributed to conflicting intentions [5] between the pilot crew and the automation.

The European project Applying Pilot Models for Safer Aircraft (A-PiMod) [6] addresses two major issues of the aviation domain: (1) poor pilot crew-automation partnership, and (2) confusion over the status of automation. As introduced above, both issues are highly connected to each other. In order to tackle these issues, the project aims to develop a virtual crew member, which takes the position of classical aircraft automation. The virtual crew member should perfectly integrate into the pilot crew resulting in a cooperative human-machine cockpit crew. As part of the crew, the virtual crew member must be able to anticipate the internal states of the human crew members, as well as the human crew members must be able to anticipate the internal states of the human and virtual crew members. This ability helps, e.g., to improve the task share in the cockpit by means of dynamic adaptions of task distributions. The concept underlying the A-PiMod project is sketched in Figure 1. A mission is achieved cooperatively by sharing tasks according to the individual capabilities of each crew member. The basis of good partnership is a sufficient understanding of each crew member about the internal states of the other crew members.

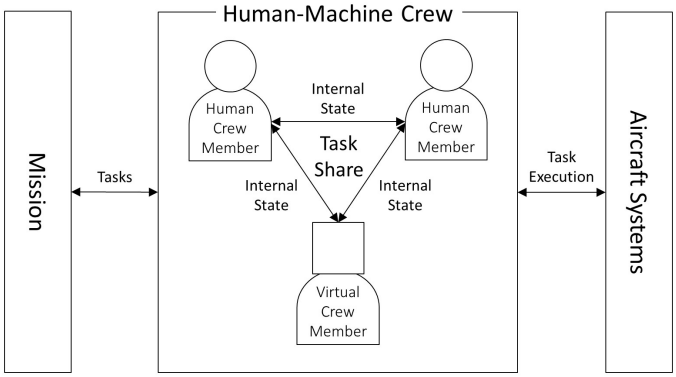


Figure 1: Partnership in the aircraft cockpit between human crew members and a virtual crew member.

In this paper, we present the concept of the A-PiMod pilot model, which will be used for inferring the internal state of the human crew members. The concept of our pilot model combines cognitive and probabilistic modelling approaches. The internal state is composed of different sub-states, which have been defined during the initial phase of the project. The addressed sub-states are situation awareness, workload, and intentions. The sub-states will be inferred based on real-time data about the mission, tasks, and pilot behaviors, including what they say, where they look at, and how they act.

In Section II, a short overview of cognitive and probabilistic operator models and some of their applications is given. Section III introduces the pilot model and describes the different target states. The integration into the A-PiMod architecture and the interaction with other A-PiMod components is explained in Section IV. Section V concludes and reveals future steps.

II. RELATED WORK

There is a great effort within the human modelling community to develop operator models to support the development of complex human-machine systems. The technology underlying

 ISBN: 978-1-61208-390-2

these models is as diverse as the purpose of using them. The A-PiMod pilot model combines cognitive and probabilistic modelling approaches. For this reason, we provide an overview of these modelling approaches in this section.

A. Cognitive Models

Cognitive models are intended to describe mental processes of human agents. An overview of extant cognitive computational models is provided in [7]–[11]. Cognitive models describes cognitive processes like human perception, decision making, memory and learning processes. When cognitive models are implemented in software, they can be used to simulate human behavior and to predict human error. For example, these cognitive architectures can be used to support the development of user interfaces in early design phases. A cognitive architecture can be understood as a generic interpreter that executes formalized task models in a psychological plausible way.

Cognitive architectures were established in the early eighties as research tools to unify psychological models of particular cognitive processes [12]. The most noted cognitive architectures are Adaptive control of Thought Rational (ACT-R) [13][14], State Operator Apply Result (SOAR) [15][16] and Man-Machine Integration Design and Analysis System (MIDAS) [17][18]. These early models only dealt with laboratory tasks in non-dynamic environments [19][20]. Furthermore, they neglected processes such as multitasking, perception and motor control that are essential for simulating human-machine interaction in highly dynamic environments. Models such as ACT-R and SOAR have been extended in this direction [21][22] but still have their main focus on processes suitable for static, non-interruptive environments. Other cognitive models like MIDAS [23], Architecture for Procedure Execution (APEX) [24] and Cognitive Network of Tasks (COGNET) [25] were explicitly motivated by the needs of human-machine interaction and thus focused for example on multitasking right from the beginning. The cognitive architecture Cognitive Architecture for Safety Critical Task Simulation (CASCaS) was developed by [26] and recognized by [27] as one of the best in the world. CASCaS has been applied in several projects, in order to analyse perception [28], attention allocation [10][29], decision making [26], and error [26][30] of humans in the automotive and aviation domains.

B. Probabilistic Models

While cognitive architectures are usually based on rules (CASCaS) or semantic networks (MIDAS) other approaches utilize probabilistic methods to model human operators. In [31] the author employs Hidden Markov Models (HMM), to describe the instrument scanning behaviour of aircraft pilots. In [32] HMMs are used to infer on the behaviour of operators of unmanned aerial vehicles and the currently performed task by monitoring operators' interactions with a User Interface.

In the automotive domain, there are approaches which employ probabilistic driver models. In [33], a hierarchical structure of Dynamic Bayesian Networks (DBN) is used to generate driving behaviours and actions from driving goals. It is also shown that it is possible to derive the behaviours and driving manoeuvers of the driver from his actions. Another example for the inference and classification of driving behaviours can be found in [34]. In the domain of Intention Recognition Systems, DBNs are used to determine the intentions of drivers [35] and to infer the intent of software users to provide specific help [36].

C. Related Applications

The pilot model used in the Crew Assistant Military Aircraft (CAMA) consists of a petri-net based part to model the normative pilot behaviour and a adaptive part which is based on fuzzy rules [37]. The adaptive part determines if deviations from the normative model are errors or were intended by the pilot due to, e.g., high workload.

Cognition Monitor (COGMON) [38] is a multidimensional approach to provide information about the state of a aircraft pilot. It relies on subjective, contextual, behavioural and physiological measures. However, to collect the physiological data intrusive techniques like electroencephalography are used, which we aim to avoid in A-PiMod.

III. Crew State Inference

Knowledge about the cognitive pilot crew state provides the possibility of adapting the systems state accordingly. The cognitive pilot crew state cannot be estimated as a whole. Instead, the cognitive pilot crew state has to be decomposed into target states which will be estimated. In the past, there has been research on a broad range of target states, such as situation awareness [39][40] and workload [41][42]. Although there may be more target states it turned out, during the requirements engineering phase of A-PiMod that the Crew State Inference (CSI) will be focused on the following target states:

- Intentions: Does the pilot intend to perform the tasks he is assigned to or is he intending to do something different?
- Workload: In how far is the pilot cognitively and physically used to capacity?
- Situation awareness: Is the pilot aware of the things around him that are of interest in context of the flight task?

For the acquisition of the necessary data for the defined target states, the Crew State Inference relies on non-intrusive techniques. The CSI infers the state of every human crew member separately. The intention is the first target state to be estimated. The results of the intention recognition are also used as inputs for the assessment of the situation awareness.

A. Intention

The A-PiMod architecture provides adaptiveness in the manner of automation and crew-automation interaction. To realize this adaptiveness the automation needs information about when to provide assistance or when it is necessary to interfere. To determine this, it is desirable for the system to know the pilot's intention. If the intention is not consistent with the situation known to the system, which could lead to a critical situation, there would be a need for further interaction with the crew.

In [43], intention is described as a composite concept specifying what the agent has chosen and how the agent is committed to that choice. The agent from this statement can be a pilot and his choice refers to a goal. This reflects that the pilot's intentions are strongly connected to the goal he is actually trying to achieve. As already mentioned, the agent

 ISBN: 978-1-61208-390-2

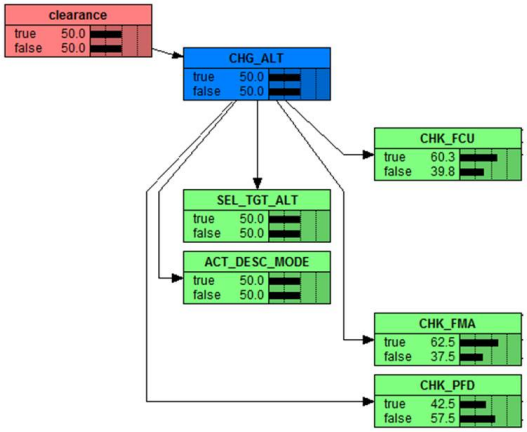


Figure 2: Basic Bayesian Network for a change altitude task

needs to be committed to this goal. That means that the agent must be able to take part in a plan, which is needed to achieve the goal. A plan is mainly a certain behaviour, which is a sequence of observable actions, that leads to the achievement of a specific goal. Plans can be more or less complex. Complex plans usually can be separated into sub-plans. Thus, the complex plans become goals of their sub-plans. In the literature the intention recognition becomes plan recognition in this case. The tasks of a pilot also serve the achievement of a goal. Complex task can be separated into sub-tasks and can become the goals of their sub-task, too. So, a task of a pilot can be interpreted as equivalent to a plan or a goal. To execute a task, a pilot has to show a specific behaviour which consists of certain actions. This means that, for each task, there exists some set of actions which is typical for this task. Many of these actions can be observed, e.g., interactions with the conventional cockpit interfaces or touch displays, or the gaze on instruments. On the basis of these observations the CSI Intention module infers on the tasks which are currently performed by the pilot. For the task inference, a Bayesian network is used. A basic example for the task to change the altitude is shown in Figure 2. The depicted network is a segment of the currently implemented network. The nodes in the network can be, depending on the type of information they represent, divided into the following groups: task nodes, context observation nodes, and action nodes. The node *CHG_ALT* represents the task, the node *clearance* is a general observation of the context and represents the availability of an clearance from Air Traffic Control (ATC) for an altitude change. Context information can make the inference more robust if the action patterns of tasks are very similar. The nodes *SEL_TGT_ALT* and *ACT_DESC_MODE* are actions and represent interactions with the Flight Control Unit (FCU). *CHK_FCU*, *CHK_PFD* and *CHK_FMA* are also actions and represent if the pilot has looked, e.g., at the FCU. Every network node has a probability table which quantifies the influence of the parent nodes on the current node. For nodes without parents (no incoming edges) a-priori probabilities have to be defined. If an action is performed, its corresponding node gains the state *true*, this results in an increasing probability of all tasks that can cause this action. The Bayesian network is currently constructed manually. The structure is based on a task analysis which was made in advance. The necessary parameter values for these probability tables are currently chosen on the basis of this task analysis. These values will be revised on the basis of the data which will be collected during simulator experiments. The intention inference delivers for each task node a probability value that this task is currently being executed by the currently considered pilot. The tasks with a probability value above a certain threshold are interpreted as the currently executed tasks of the pilot. These are the subjective tasks of a pilot, which are the output of the intention inference module. Thanks to this approach based on a Bayesian network, we will be in the position to recognize the pilot's intentions seen as goals or tasks.

B. Workload

High, as well as too low workload can influence the pilots' performance negatively. Thus, the purpose of the workload module is to determine in how far the pilot is cognitively and physically used to capacity. In this module the workload, of a pilot is described by a multi-resource model which is comparable to the one of Wickens [44]. The dimensions of our workload model are Visual perception, Visual processing, Auditory perception, Auditory processing, and Auditory action. According to this model the pilots' cognitive and physical capacities in the different dimensions are limited. The execution of tasks causes the consumption of some of these capacities, which leads to an increased workload. The relevant tasks were identified and described during a task analysis. The description of every identified task is stored in a task pool. The task description contains, among other things, information about the workload which is created by a task in the different dimension. These workload values were collected by interviewing pilots with a questionnaire. To estimate the current workload of the pilot, his objective tasks, the tasks he is currently assigned to are taken into account. The workload values stored in the task descriptions for each dimension are summed up over the pilot's objective tasks. These aggregated workload scores reflect the actual load of the pilot in the different dimensions and are the output of the workload module.

C. Situation Awareness

Situation awareness (SA) is a state of knowledge which is the product of a cognitive process, called situation assessment. During situation assessment, operators interpret available environment and system information in the context of their current goals. Based on SA, operators decide what they are going to do in a certain situation. Due to high automation, loss of SA can remain without bad consequences in standard situations but can lead to accidents in critical situations. Knowing the current coverage of SA at each moment in time during operation could help to prevent incidents and accidents caused by incorrect SA. The purpose of the SA module of the CSI is real-time SA inference. The basis for the inference process is a formal situation model and a formal model of SA which represents the human operators subjective state. Both models consist of a set of atomic elements. These elements can refer to basic items like information or more complex things like tasks. Information elements are linked to the perception level of the SA while the tasks are more related to the comprehension level.

 ISBN: 978-1-61208-390-2

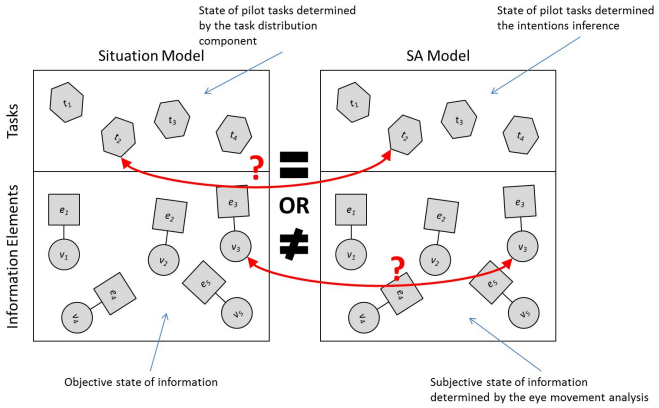


Figure 3: Description of SA inference for perceptive aspects

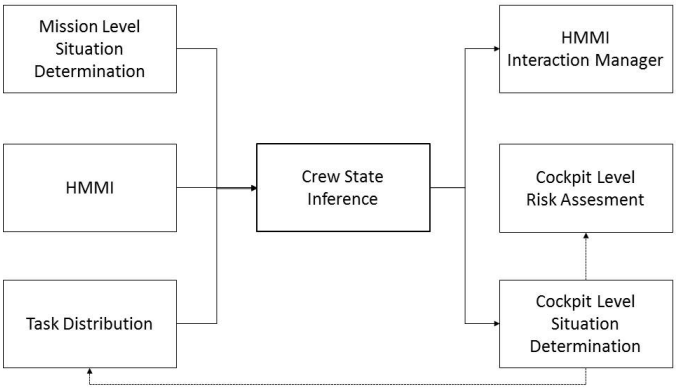


Figure 4: Crew State Inference module connections to other modules inside the A-PiMod architecture (connections between most of the other modules were intentionally left out)

The human operator can update the subjective model by, e.g., focusing on sources of information. The state of each atomic element of the SA model can be compared to the state of each element of the situation model. The approach is visualized in Figure 3. The comparison allows detecting inconsistencies between the real situations and situations as they are perceived by the operator. Currently, SA is focused on the tasks of a human operator. The tasks of the Situation Model are the objective tasks of the pilot. The current set of objective tasks is delivered by an external component, the Task Distribution module. The active subjective tasks of the SA model are updated by the Intention inference module. The objective and the subjective tasks of the human operator are compared. Thus it can be determined if the operator performs the tasks which are appropriate for the current situation. With an eye-tracker it would also be possible to consider the basic information elements. These elements are then updated in the subjective model whenever the pilot gazes on the corresponding cockpit instrument.

IV. Integration into A-PiMod Architecture

The A-PiMod architecture consists of several new components which do not exist in present cockpits. The aim is to provide further assistance and to improve the interaction between the human pilots and the automation. The new components are Mission Level Situation Determination, Mission Level Risk Assessment, Cockpit Level Situation Determination, Task Determination at Cockpit Level, Task Distribution, Cockpit Level Risk Assessment, Human Machine Multimodal Interface (HMMI), HMMI Interaction Manager and Crew State Inference. The Crew State Inference communicates with the components as shown in Figure 4. Input data is received from the Mission Level Situation Determination, the HMMI and the Task Distribution. The output of the CSI module is aggregated by the Cockpit Level Situation Determination and then processed by the Cockpit Level Risk Assessment and the Task Distribution. The CSI also provides input for the HMMI Interaction Manager

Mission Level Situation Determination delivers context information like the progress on flight plan, the state of aircraft systems and environmental data (e.g., weather, ATC). The data of this component are treated as general observations of context in the CSI intention inference module.

The HMMI handles interactions of the human crew members with the cockpit in several modalities. The supported modalities are conventional buttons, touch, speech and gestures. Additionally, this component tracks the eye movements of the human crew members. Thus, HMMI delivers the actions and the gaze information of the pilots which are interpreted by the CSI.

The Task Distribution component receives, from the Task Determination at Cockpit Level, the tasks which are pertinent for the given situation. Every pertinent task is then assigned to at least one crew member (including automation) which is capable of performing this task. To elaborate a new task distribution the component considers the capabilities and the state of all available crew members, including human pilots and automation systems. The currently active task distribution is communicated to the CSI and consists of a set of tasks for each crew member. The tasks of a set are the so-called objective tasks of a human operator.

The Cockpit Level Situation Determination aggregates the states of all crew members. This means it monitors the state of all automation systems and receives the state of the human pilots from the CSI component. This information is then delivered to the Task Distribution component and the Cockpit Level Risk Assessment component.

Cockpit Level Risk Assessment evaluates the risk for task distributions. Here, the state of the human crew members and the state of the automation system is taken into account. Only if the risk for a task distribution is below a certain threshold this task distribution can be activated. If there are more than one possible task distribution available usually the one with the lowest risk is activated.

The purpose of the HMMI Interaction Manager is to modify the salience and the modality of the HMMI output to the human crew members. This means, e.g., if a human crew member is not aware of some information, the Interaction Manager can make the information on the display more salient. If the human crew member currently has a high visual workload, the Interaction Manager could also switch the output modality of the information to speech.

 ISBN: 978-1-61208-390-2

V. CONCLUSION

In this paper, we presented the concept of the A-PiMod pilot model, which uses a mixed modelling approach (cognitive and probabilistic) to infer intentions, situation awareness, and workload of a pilot crew. The pilot model is embedded into the A-PiMod architecture and relies on the data generated by other modules. A first prototype of the pilot model has been integrated and the communication with other modules has been tested within a simulator setting at the German Aerospace Center. The next steps will be to improve the concept and the implementation, in order to be ready for a first validation of the promised functions.

ACKNOWLEDGMENT

The A-PiMod project is funded by the European Commission Seventh Framework Programme (FP7/2007-2013) under contract number: 605141 Project A-PiMod.

REFERENCES

[1] R. Amalberti, "Automation in aviation: A human factors perspective," in Handbook of aviation human factors. Lawrence Erlbaum Associates Mahwah, NJ, 1999, pp. 173–192.

[2] R. Parasuraman and V. Riley, "Humans and automation: Use, misuse, disuse, abuse," in Human Factors: The Journal of the Human Factors and Ergonomics Society, vol. 39, no. 2. SAGE Publications, 1997, pp. 230–253.

[3] R. Parasuraman and C. Wickens, "Humans: Still vital after all these years of automation," in Human Factors: The Journal of the Human Factors and Ergonomics Society, vol. 50, no. 3. Sage Publications, 2008, pp. 511–520.

[4] L. Bainbridge, "Ironies of automation," in Automatica, vol. 19, no. 6. Elsevier, 1983, pp. 775–779.

[5] H. Sogame and P. Ladkin, "Aircraft accident investigation report 96-5. japan: Ministry of transport," 1996. [Online]. Available: http://sunnyday.mit.edu/accidents/nag-1.html [retrieved: 1,2015]

[6] "Applying pilot models for safer aircraft." [Online]. Available: http://www.apimod.eu [retrieved: 2,2015]

[7] A. Mavor et al., Modeling human and organizational behavior: Application to military simulations. National Academies Press, 1998.

[8] K. Leiden et al., "A review of human performance models for the prediction of human error," in Ann Arbor, vol. 1001, 2001, p. 48105.

[9] J. Rasmussen, "Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models," in Systems, Man and Cybernetics, IEEE Transactions on, vol. 12, no. 3. IEEE, 1983, pp. 257–266.

[10] F. Frische, J.-P. Osterloh, and A. Lüdtke, "Modelling and validating pilots visual attention allocation during the interaction with an advanced flight management system," in Human Modelling in Assisted Transportation. Springer, 2011, pp. 165–172.

[11] F. E. Ritter et al., "Techniques for modeling human performance in synthetic environments: A supplementary review," DTIC Document, Tech. Rep., 2003.

[12] A. Newell, Unified theories of cognition. Harvard University Press, 1994.

[13] J. R. Anderson and C. Lebiere, The atomic components of thought. Psychology Press, 1998.

[14] J. R. Anderson, How can the human mind occur in the physical universe? Oxford University Press, 2007.

[15] A. Newell and H. A. Simon, GPS, a program that simulates human thought. Defense Technical Information Center, 1961.

[16] J. F. Lehman, J. E. Laird, and P. S. Rosenbloom, "A gentle introduction to Soar, an architecture for human cognition," in Invitation to cognitive science, vol. 4. MIT Press, 1996, pp. 212–249.

[17] K. M. Corker and B. R. Smith, "An architecture and model for cognitive engineering simulation analysis: Application to advanced aviation automation," in Proceedings of the AIAA Computing in Aerospace 9 Conference, 1993, pp. 1079–1088.

[18] B. F. Gore, "Workload as a Performance Shaping Factor for Human Performance Models," in Behavioral Representation in Modeling and Simulation (BRIMS), 2011, p. 276.

[19] J. R. Anderson, Rules of the mind. Psychology Press, 2014.

[20] A. Newell, P. S. Rosenbloom, and J. E. Laird, "Symbolic architectures for cognition," DTIC Document, Tech. Rep., 1989.

[21] J. R. Anderson et al., "An integrated theory of the mind." in Psychological review, vol. 111, no. 4. American Psychological Association, 2004, p. 1036.

[22] R. E. Wray and R. M. Jones, "Considering Soar as an agent architecture," in Cognition and multi-agent interaction: From cognitive modeling to social simulation, vol. 33, 2006, pp. 53–78.

[23] K. M. Corker, "Cognitive models and control: Human and system dynamics in advanced airspace operations," in Cognitive engineering in the aviation domain, vol. 31. Lawrence Erlbaum Associates, 2000, pp. 13–42.

[24] M. A. Freed, "Simulating human performance in complex, dynamic environments," Ph.D. dissertation, Northwestern University, 1998.

[25] W. Zachary, T. Santarelli, J. Ryder, and J. Stokes, "Developing a multi-tasking cognitive agent using the COGNET/iGEN integrative architecture," DTIC Document, Tech. Rep., 2000.

[26] A. Lüdtke, L. Weber, J.-P. Osterloh, and B. Wortelen, "Modeling Pilot and Driver Behavior for Human Error Simulation," in Digital Human Modeling, ser. Lecture Notes in Computer Science, V. Duffy, Ed., vol. 5620. Springer, 2009, pp. 403–412. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-02809-0_43 [retrieved: 1,2015]

[27] C. Wickens et al., "Modeling and evaluating pilot performance in nextgen: Review of and recommendations regarding pilot modeling efforts, architectures, and validation studies," NASA Ames Research Center, Moffett Field, CA, Tech. Rep. NASA/TM-2013-216504, 2013.

[28] A. Lüdtke and J.-P. Osterloh, "Simulating perceptive processes of pilots to support system design," in Human-Computer Interaction–INTERACT 2009. Springer, 2009, pp. 471–484.

[29] B. Wortelen, A. Lüdtke, and M. Baumann, "Integrated simulation of attention distribution and driving behavior," in Proceedings of the 22nd Annual Conference on Behavior Representation in Modeling & Simulation, W. G. Kennedy, R. S. Amant, and D. Reitter, Eds. Ottawa, Canada: BRIMS Society, 2013, pp. 69–76.

[30] A. Lüdtke, J.-P. Osterloh, T. Mioch, F. Rister, and R. Looije, "Cognitive modelling of pilot errors and error recovery in flight management tasks," in Human Error, Safety and Systems Development. Springer, 2010, pp. 54–67.

[31] M. Hayashi, "Hidden Markov Models for analysis of pilot instrument scanning and attention switching," Ph.D. dissertation, Massachusetts Institute of Technology, 2004.

[32] D. Donath, "Verhaltensanalyse der Beanspruchung des Operateurs in der Multi-UAV-Führung," Dissertation, Universität der Bundeswehr München, 2012.

[33] C. Moebus and M. Eilers, "Prototyping Smart Assistance with Bayesian Autonomous Driver Models," in Handbook of Research on Ambient Intelligence and Smart Environments: Trends and Perspectives, N.-Y. Chong and F. Mastrogiovanni, Eds. IGI Global, May 2011, pp. 460–512. [Online]. Available: http://www.igi-global.com/chapter/prototyping-smart-assistance-bayesian-autonomous/54671 [retrieved: 1,2015]

[34] G. Agamennoni, J. I. Nieto, and E. M. Nebot, "A bayesian approach for driving behavior inference," in 2011 IEEE Intelligent Vehicles Symposium (IV), no. Iv. Ieee, 2011, pp. 595–600.

[35] M.-I. Toma and D. Datcu, "Determining car driver interaction intent through analysis of behavior patterns," in Technological Innovation for Value Creation. Springer, 2012, pp. 113–120.

[36] E. Horvitz, J. Breese, D. Heckerman, D. Hovel, and K. Rommelse, "The Lumiere project: Bayesian user modeling for inferring the goals and needs of software users," in Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence. Morgan Kaufmann Publishers Inc., 1998, pp. 256–265. [Online]. Available: http://dl.acm.org/citation.cfm?id=2074124 [retrieved: 1,2015]

[37] M. Strohal and R. Onken, "Intent and error recognition as part of a knowledge-based cockpit assistant," in Proc. SPIE, vol. 3390, 1998,

 ISBN: 978-1-61208-390-2

pp. 287–299. [Online]. Available: http://dx.doi.org/10.1117/12.304818 [retrieved: 1,2015]

[38] C. W. Pleydell-Pearce, B. Dickson, and S. Whitecross, "Cognition monitor: a system for real time pilot state assessment," in Contemporary Ergonomics. Taylor & Francis Group, 2000, pp. 65 – 69. [Online]. Available: http://research-information.bristol.ac.uk/en/publications/cognition-monitor-a-system-for-real-time-pilot-state-assessment(fbd8abcd-d97e-4963-b042-c8a5e4a5f5dd).html [retrieved: 1,2015]

[39] K. S. Moore, "Comparison of Eye Movement Data to Direct Measures of Situation Awareness for Development of a Novel Measurement Technique in Dynamic, Uncontrolled Test Environments," Ph.D. dissertation, Clemson University, 2009.

[40] M. Diez et al., "Tracking pilot interactions with flight management systems through eye movements," in Proceedings of the 11th International Symposium on Aviation Psychology, 2001, pp. 1–6.

[41] D. Donath and A. Schulte, "Behavior Model Based Recognition of Critical Pilot Workload as Trigger for Cognitive Operator Assistance," in Engineering Psychology and Cognitive Ergonomics. Springer, 2009, pp. 518–528.

[42] T. C. Hankins and G. F. Wilson, "A comparison of heart rate, eye activity, EEG and subjective measures of pilot mental workload during flight." in Aviation, Space, and Environmental Medicine, vol. 69, no. 4, 1998, pp. 360–367.

[43] P. R. Cohen and H. J. Levesque, "Intention is choice with commitment," in Artificial Intelligence, vol. 42, no. 2-3, Mar. 1990, pp. 213–261. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/0004370290900555 [retrieved: 1,2015]

[44] C. D. Wickens, "Processing Resources in Attention," in Varieties of attention. Academic Press, 1984, pp. 63–102.

 ISBN: 978-1-61208-390-2