TouchWear: Context-Dependent and Self-Learning Personal Speech Assistant for Wearable Systems with Deep Neural Networks

Using Contextual LSTMs on Recurrent Neural Networks

Joshua Ho
Research Center for Information Technology Innovation
Academia Sinica
Taipei, Taiwan
jho@iis.sinica.edu.tw

Chien-Min Wang
Institute of Information Science
Academia Sinica
Taipei, Taiwan
cmwang@iis.sinica.edu.tw

***Abstract*— Context awareness in future adaptive systems for wearable computers comprise many features, such as ability to sense and perceive contexts, to be inferred by the generation of a user model, to perform the computation and present the communication interface, and to provision implemented services. In this work, we introduce a system application prototype implemented by distinguishing contexts from wearable systems. Thus, user behavior, activity, and application data are trained to generate a user model. Next, a voice interface administered by the artificial personal speech assistant not only enables conversation with the user but is also used to build a recurrent model of deep neural networks primarily based on the conversation logs. Ultimately, the service and recommendation framework are implemented and deployed so that the wearable system has the capacity to aid people in need by means of service-oriented and wearable adaptation.**

Keywords-wearable computing; personal speech assistant; context awareness; deep neural network.

I. INTRODUCTION

Using wearable devices, like smart watch and smart glasses, is more than just convenient, because they collect important information about the context according to the user's body behavior and head movement. In contrast to smartphone users who often receive limited information of little on-body context because of their 'heads-down' gesture from looking at their smartphones, users of wearable devices are able to focus more on social interactions and the surrounding views. Wearing smart watches and smart glasses permits fewer restrictions and more augmented conditions.

In recent years, artificial speech assistants like Apple Siri, Amazon Alexa, Cortana, and Google Assistant [19] have become widely adopted as a conversing medium for mobile computation. By taking advantage of acoustic and concatenative models of Text-To-Speech (TTS), speech assistants can execute and control voice commands, system recommendations and services according to user requests. In our design, recommendations and services can more effectively conform to personal intentions, activities, favorites, records, surrounding environments, social networks and crowdsourced information.

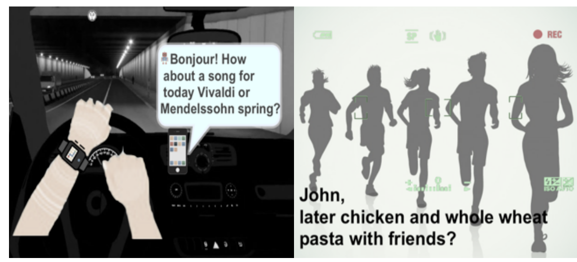


Figure 1. Context dependent states with wearable systems are automatically learned and identified by the Personal Speech Assistant of DNN, which continously offers relevant and appropriate services.

Therefore, the proposed system, TouchWear, aims to use wearable computers to present contextual, automated sensing as well as a service-oriented workflow for human computation (Figure 1). Furthermore, TouchWear is represented by the personal speech assistant (PSA) that models with continuous self-learning Deep Neural Networks (DNN), to transform and retrieve helpful, on-the-fly, historical, or even private information. Finally, the system provides mobile services that hinge on the infrastructure being successfully used in practical deployment.

This paper introduces the implementation of our prototype application for wearable devices, which is built on context dependent and continuous information from the user perspective based on modern Artificial Intelligence of DNNs. TouchWear is designed according to the following four methods: (1) integration of previous research paradigms for recognizing user activity into the system, and the design of a system adapter for context awareness through the use of wearable devices; (2) evaluation of learning patterns from a user's behavior and a conversation proposed by the PSA, and performance of continuously self-learning AI model based on DNNs, where the system adapter is flexible enough to manage either notifying the PSA of recognized contexts, or perceiving new contexts; (3) design of a message extractor and filter to better address the user's contextual query, while at the same time, personalized results retrieved and generated by the PSA processed every now and then; (4) implementation of an infrastructure for mobile service-oriented applications (SOAs), which model the business requirement and bring services via specially designed user interfaces.

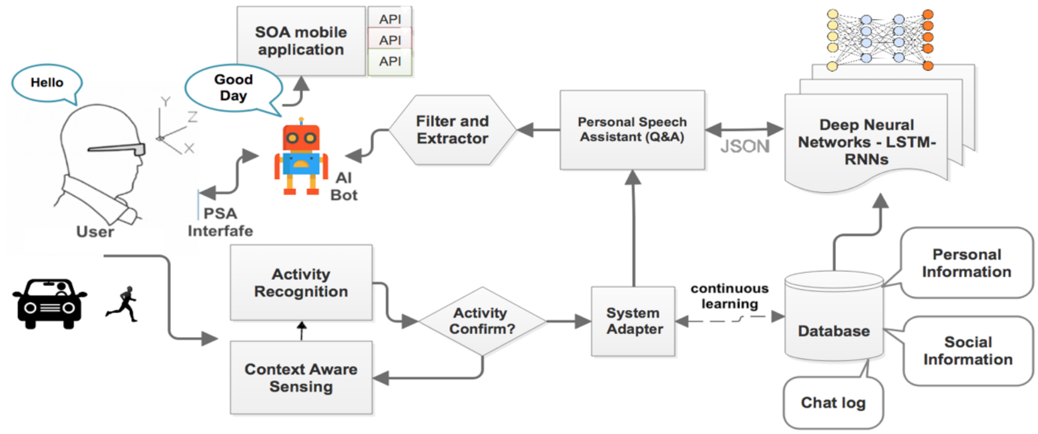


Figure 2. TouchWear system architecture – an overview of wearable application.

The organization of this paper is presented as follows: Section II provides a review of the previously related works. Section III depicts the design of the system architecture and its challenges. Section IV evaluates the use cases and preliminary results, and finally, Section V concludes the paper with our current plan for the future work.

II. RELATED WORK

In recent years, research has discussed many areas that are related to our work, such as Human Activity Recognition (HAR) [1][5][6] on mobile devices, which helps us understand how to analyze user's wearable sensory systems and the designed interfaces [2][3] for perceiving contexts. Context-awareness involves the concept of sensing oneself in a context, which means tracking 'Head-Centered and Context-Aware Learning' [12][15][16] on a wearable device, and also exploring the surrounding environments. Pervasive computing to address location-awareness [4] has also drawn considerable attentions in the development of wearable computers [7]. Other related work on wearable devices tackles privacy-preserving issues [11] on a crowd-powered system, which also inspires us for designing and enhancing our information retrieval, filtering, and extraction.

Furthermore, conversational agents with Artificial Intelligence are becoming increasingly ubiquitous in business, technology and daily life. Relevant research on these agents describe PSA to recognize the disordered speech [18], Chatbots with a Support Vector Machine (SVM) classifier [9], end-to-end systems [21][22] to play the communication role to synchronize physical motoring [10], and DNN-based agents to build embedded questions and answers, based on bidirectional long short-term memory (LSTM) network to measure the cosine similarity [8][17].

In order to achieve ubiquitous data access on mobile and wearable computing in TouchWear, SOAs are practiced and designed due to the limited memory and connection bandwidth [13]. Based on the advocated services designed and implemented by SOAs [14], our proposed system is able to consider user's adaptive contexts as predicted services via PSA more adequately and efficiently than the related works.

III. SYSTEM ARCHITECTURE

The system guides a user through the designed wearable application (Figure 2), while the PSA provides instructions and conversations on the voice-based application. The below steps present the processes, integrated frameworks, components and how they work together.

A. Context Aware Sensing and Wearable Devices

Contextual sensing is the most fundamental analysis of context-aware systems. TouchWear directly uses sensory data of Accelerometer, Gyroscope, and the signals of Global Positioning Systems (GPS) to detect a user's activity and location, where Wi-Fi signals are also considered in the indoors [24]. With our wearable devices (Google Glass [25], Sony SmartEyeglass [26]) and producing data (frequency 5Hz), the modeled SVM classifier is capable of recognizing targeted activity and location in around 3 seconds.

B. Activity Recognition and System Adapter

In Table 1, we depict six activities (both indoors and outdoors) in which TouchWear takes the detected context-aware messages as prerequisite information to prepare for the conversations with the user. The system adapter, which is based on context awareness, will inform PSA per user's request to initiate the conversation. As for the content of the conversation, the DNN will periodically notify the PSA via APIs if there is any update to the latest entropy. Furthermore, continuous self-learning occurs to conceive new contexts, such as new activities, or to improve accuracy of old ones.

APIs can be triggered by the following: highly compressed formats, publishing and exchanging protocols, Web Services with SOAP, XML-based service invocation, JSON RESTful services implementing TouchWear, and interface compliance with Open Standard Gateway initiative (OSGi). The designed adapter needs not only to implement the regulations satisfying the requirement of each application, but also to use the exact pair of enterprise public and private key infrastructure (PKI), SSL, or the secure PGP encryption system [20] to cryptographically achieve needs.

TABLE I. RECOGNIZED ACTIVITY AND LOCATIONS

Smart Glasses Activity, Location, and Performance			
Activity	***Outdoor***	***Indoor***	***Performance***
driving	city road, highway	N/A	87%
jogging	hiking route, mountain area	gym, indoor stadium	86%
walking	side walk, street	building hallway, house	88%
sitting	outdoor bench, park, open field	office, study room, living room	89%
cooking	BBQ, brewery area	kitchen, dining room	87%
dining	places for grilling, garden	dining room, restaurant	91%

C. *Database and Deep Neural Networks*

Early works on computer speech systems focused on rule-based or hand-crafted implementations to simulate human conversations [9]. However, it is very difficult to enumerate the real conditions and all possible states, especially in light of the great complexity of human language. For this reason, recent speech assistants and Chatbots in Recurrent Neural Networks (RNNs) of DNN have been shown to meliorate accuracy to improve performance. After the evaluations of two large datasets, the Cornell Corpus of movie dialogues and thousands of Twitter logs with Long Short-Term Memory networks (LSTMs), TouchWear takes sequence to sequence (seq2seq) learning process [23] to construct its memory dependent network by using the conversation logs at this stage. Information of personal (such as emails) and social (Emails, tweets) is stored in the database server, and also continuously migrated to model the recurrent contextual information in the proposed system.

D. *Personal Speech Assistant*

The PSA, or AI Bot in TouchWear, is implemented by using the open-source project of Google Hangouts [30] to leverage current applications on our wearable platform. The AI Bot uses contextual information of activity recognition according to the wearable application; then, the AI Bot initials the automated service with example greetings such as "*Hi buddy, would you like some music while driving?*" or "*Good morning John, how may I help you?*", which are the first contextual messages. In contrast to these examples in which the AI Bot initiates the conversation, users for instance can just simply say "*Please mute, Bot*" to switch it back to the on-demand service type. So, "*OK, AI Bot*" or "*Hi Bot*" are launched by user to converse with the AI Bot.

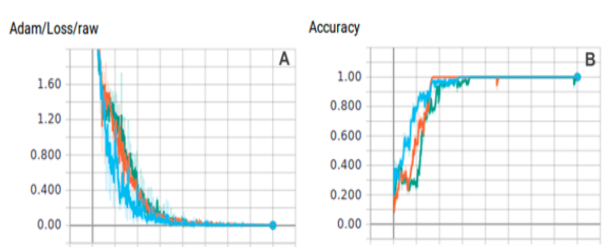


Figure 3. (A) Loss, (B) Accuracy: Training after 1000 epochs, where three testing datasets were evaluated.

There were three testing datasets that contain conversation logs trained by the LSTMs in our DNN. Since the seq2seq training processes use the same training data to validate the model in each epoch done by TensorFlow [27] and tflearn [28] frameworks, an iteration of 1000 epochs generates a loss of 0.00385 and an approaching accuracy of 1.00000 near the 400th epoch visualized on the *tensorboard* [27] (Figure 3).

E. *Filter and Extractor*

The retrieved responses are provided by AI Bot according to the deep learning from ongoing conversations, user Email threads, and simulated social tweets. TouchWear currently uses four categories to filter and extract the results, as shown in Table II. The four categorized directories in the system are *regular* and *critical* for *personal information*, and *regular* and *privacy preserving* for *social information*. Accordingly, the authentication plays a vital role in the '*critical*' category for *personal information*, whereas filtered datasets, using metadata and programming, are particularly essential for the '*privacy preserving*' category for *social information*.

TABLE II. INFORMATION FILTER AND EXTRACTOR

Information Type	**Category vs. Filter and Extractor**		
	Category	***Filter***	***Extractor***
personal information	*regular*	none	ranking
	critical	authenticated by secure frameworks	extracted ranking based on secure frameworks
social informaion	*regular*	defined rules	ranking
	privacy preserving	filtered datasets	extracted ranking based on filtered datasets

F. *Service-Oriented Mobile Application*

Mobile SOAs are examined and designed for TouchWear. The backend servers receive user commands through the PSA, and the commands are executed by contracting the system APIs of the targeted application. If the syntax is complying with the regulations and if the user's authentications are authorized, the provisioning applications will be triggered and planned toward the completion to meet business requirements. At the present time, the system has 7 mini services (or groups) to evaluate the system integrity in the experimental and validating phase. Table III below shows the list of SOA mini-services, where the services with asterisk have the permission to access the personal contacts.

TABLE III. SOA MINI SERVICE LIST

SOA Services
1. Voice or video call *
2. Search and play music (personal music albums)
3. Facility automation
4. Search 'keyword': conversation logs, Email *
5. Social networks: recommendation for shopping and entertainment *
6. GPS navigation setup
7. Food / restaurant search and reference

 ISBN: 978-1-61208-657-6

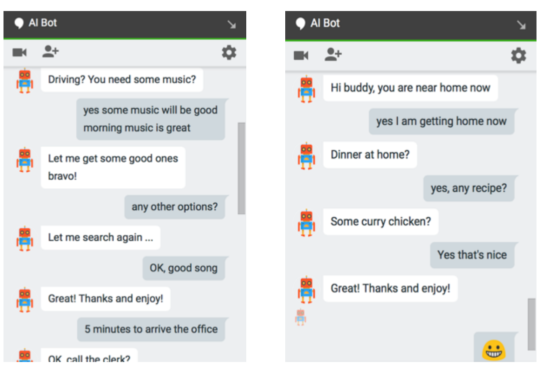
G. *Other System Frameworks*

The system stands on the top of TensorFlow to build up DNNs with LSTM RNN, which is implemented by seq2seq learning process. In seq2seq, the encoder and decoder take the input and generate the output based on the semantic contexts. In our experiments, we observed that LSTM could learn to spell words and copy general syntactic structures to capture the essence of the input sentences. Thus, the system was prepared with initial trials of training data that consisted of 1025 Email threads, 480 lines of conversation log on Hangouts, and dozens of social tweets simulated by the open-source SocialEngine [29].

IV. USE CASES AND PRELIMITARY RESULTS

We began our project with the aim of studying how to recognize activity and location by using sensors on wearable devices. The current system can recognize three activities, *driving*, *jogging*, *walking* with an accuracy up to 87% and its performance is getting better in our experiments. However, though *sitting* can be recognized with accuracy 89%, it is more difficult to distinguish *dining* and *sitting* since both are very similar, unless additional sensors like the camera and Wi-Fi signals are applied, same as for *cooking* as well. For information extraction and filtering, the current system ranks results with descending score and/or reverse chronicle order, and the top one will return to the query each time.

Initial trials of use cases were conducted using 11 types (omitting indoor *driving*) according to 6 indoor and outdoor activities that were recognized by the system. From the user's perspective, contextual services and conversing accuracy are the most important parts. Two use cases are demonstrated in Figure 4, where (a) a user is heading to work by *driving* and was successfully recommended an enjoyable song, and (b) a user is walking and located close to home or is on the way home, and the recipe recommendation of dinner is offered by AI Bot.



(a) (b)

Figure 4. Demonstrated use cases (a) driving, (b) walking.

TABLE IV. SURVEY AND POSITIVE RESPONSES

Interview Questions	Positive Responses
Q1. Are context-aware PSAs more perceived and helpful?	87.5%
Q2. How is the performance by using contextual PSA with DNNs and SOA?	83.3%

Also, the filtered datasets are designed to authenticate users before accessing their personal information so as to protect their privacy, where defined rules are given to control sharing and prevent the leaking of private information. Shared topics include food, entertainment, shopping experiences in Emails and tweets, with the removal of critical data according to filtered datasets. The initial trials were conducted in a proof-of-concept system, and the results show that the performance is very high regarding context-awareness, the conversation accuracy of LSTMs and the targeted SOAs in the laboratory, though more calibrations to our system are still further required, such as '*machine-learning search*', '*social sharing*' and '*location precision*'.

The survey from our user study with 16 participants shows that the proposed system was more preferred than systems without context-awareness (Table IV): (*i*) the wearable platform with context-aware PSAs were found to be more advantageous, as helpful aids and with more perceived accuracy; (*ii*) the performance of the contextual PSAs was seen as more resembling a real and constructive intelligent agent that assists people in their daily life. These PSAs were based on the security and the preservation of privacy of personal and social information on LSTMs, and the designed SOA mobile applications also offer contextual services per user's requests.

V. CONCLUSION AND FUTURE WORK

TouchWear proposes a unique system design for the proposed wearable application that exploits the contextual information for future wearable systems, by integrating a wearable platform, context-aware computing, PSAs, and modeling and modification of DNNs with recurrent neural networks, with the aim to design more intelligent solutions for problems that emerge in daily living. Moreover, the system is tailored to user-centric requirements and services effectively extracted by the designated information retrieval. Likewise, service operations are explicitly performed by the SOA-based mini services of mobile applications. Compared to systems without contexts, the proposed contextual DNNs significantly outperform the accuracy of conversation exchanges, start automated workflows for predicting and comprehending user's status, and take into account user favorites, demands and social associations by using the AI Bot more intimately. The continuous self-learning processes are clearly able to achieve more system genuineness, usability and user-friendliness. Our implementation was also more favored by the users according to the interviews. The insightful design of the application is promising and can be extended to the benefit of many people, their workplaces, and homes. If this forthcoming system is extensively adopted, we anticipate in the future that optimal context-awareness and context-intelligent wearable computing will be achieved in addition to Artificial Intelligence.

In our future work, we will focus other subsets of DNNs for any domain, investigate more use cases of search and social application, and identify and design more scenarios for disability-oriented systems in wearable computing. We hope our upcoming systems will assist more people in their daily living and activities.

REFERENCES

[1] Lara, Oscar D., and Miguel A. Labrador. “A Survey on Human Activity Recognition Using Wearable Sensors.” *IEEE Communications Surveys and Tutorials*, vol. 15, no. 3, 2013, pp. 1192–1209.

[2] Patel, Shyamal, et al. “A Review of Wearable Sensors and Systems with Application in Rehabilitation.” *Journal of Neuroengineering and Rehabilitation*, vol. 9, no. 1, 2012, pp. 21–21.

[3] Mistry, Pranav, et al. “WUW - Wear Ur World: A Wearable Gestural Interface.” *CHI ’09 Extended Abstracts on Human Factors in Computing Systems*, 2009, pp. 4111–4116.

[4] Ashbrook, Daniel, and Thad Starner. “Using GPS to Learn Significant Locations and Predict Movement across Multiple Users.” *Ubiquitous Computing*, vol. 7, no. 5, 2003, pp. 275–286.

[5] Lee, Seon-Woo, and Kenji Mase. “Activity and Location Recognition Using Wearable Sensors.” *IEEE Pervasive Computing*, vol. 1, no. 3, 2002, pp. 24–32.

[6] Ho, Joshua, and Chien-Min Wang. “User-Centric and Real-Time Activity Recognition Using Smart Glasses.” *GPC*, 2016, pp. 196–210.

[7] Pascoe, Jason. “Adding Generic Contextual Capabilities to Wearable Computers.” *Digest of Papers. Second International Symposium on Wearable Computers (Cat. No.98EX215)*, 1998, pp. 92–99.

[8] Tan, Ming, et al. “LSTM-Based Deep Learning Models for Non-Factoid Answer Selection.” *ArXiv Preprint ArXiv:1511.04108*, 2015.

[9] Joelianto, Endra, and Bowo Prakoso. “Support Vector Machine Classifier Based on Approximate Entropy Metric for Chatbot Text-Based Communication.” *International Journal of Artificial Intelligence*, vol. 15, no. 1, 2017, pp. 1–16.

[10] Johnson, Alex, et al. "Implementing Physical Capabilities for an Existing Chatbot by Using a Repurposed Animatronic to Synchronize Motor Positioning with Speech." *International Journal of Advanced Studies in Computers, Science and Engineering* 6.1 (2017): 20.

[11] Swaminathan, Saiganesh, et al. “WearMail: On-the-Go Access to Information in Your Email with a Privacy-Preserving Human Computation Workflow.” *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, 2017, pp. 807–815.

[12] Krause, Andreas, et al. “Unsupervised, Dynamic Identification of Physiological and Activity Context in Wearable Computing.” *Seventh IEEE International Symposium on Wearable Computers, 2003. Proceedings.*, 2003, pp. 88–97.

[13] Natchetoi, Yuri, et al. “Service-Oriented Architecture for Mobile Applications.” *Proceedings of the 1st International Workshop on Software Architectures and Mobility*, 2008, pp. 27–32.

[14] Taktak, Hajer, and Faouzi Moussa. “A Service-Oriented Application Creation Process in Ubiquitous Environments: Travel Assistant Mobile Application.” *International Journal of Pervasive Computing and Communications*, vol. 13, no. 3, 2017, pp. 300–330.

[15] Kuhn, Jochen, et al. “GPhysics—Using Smart Glasses for Head-Centered, Context-Aware Learning in Physics Experiments.” *IEEE Transactions on Learning Technologies*, vol. 9, no. 4, 2016, pp. 304–317.

[16] Kim, Ki Joon, and Dong-Hee Shin. "An acceptance model for smart watches: Implications for the adoption of future wearable technology." *Internet Research* 25.4 (2015): 527-541.

[17] Kandasamy, Kirthevasan, et al. “Batch Policy Gradient Methods for Improving Neural Conversation Models.” *International Conference on Learning Representations*, 2017.

[18] Cavalcante, Agnieszka Bętkowska, and Monika Grajzer. “Mobile and Personal Speech Assistant for the Recognition of Disordered Speech.” *SPWID 2016, The Second International Conference on Smart Portable, Wearable, Implantable and Disability-Oriented Devices and Systems*, 2016, pp. 6–10.

[19] López, Gustavo, et al. “Alexa vs. Siri vs. Cortana vs. Google Assistant: A Comparison of Speech-Based Natural User Interfaces.” *International Conference on Applied Human Factors and Ergonomics*, 2017, pp. 241–250.

[20] Hankerson D. Hernandez J.L. Kirkup M.Menezes A. Brown M., Cheung D. PGP in constrained wireless devices. *In 9th USENIX Security Symposium*., 2000.

[21] Yang, X., Chen, Y.-N., Hakkani-Tur, D., Crook, P., Li, X., Gao, J., and Deng, L. (2016). End-to-End Joint Learning of Natural Language Understanding and Dialogue Manager. *ArXiv e-prints*.

[22] Ali Orkan Bayer, Evgeny A. Stepanov, and Giuseppe Riccardi. Towards end-to-end spoken dialogue systems with turn embeddings. In *Annual Conference of the International Speech Communication Association (INTERSPEECH)*, Stockholm, Sweden, August 2017. ISCA.

[23] Sutskever, I. Vinyals, O. & Le. Q. V. Sequence to sequence learning with neural networks. In *Proc. Advances in Neural Information Processing Systems 27* 3104–3112 (2014).

[24] FIND3, “FIND3,” https://github.com/schollz/find3, July 16, 2018.

[25] Glass Explorer Edition, “Glass Explorer Edition,” https://developers.google.com/glass/, July 16, 2018.

[26] Sony Smarteyeglass SED-E1, “Sony Smarteyeglass SED-E1,” https://developer.sony.com/develop/smarteyeglass-sed-e1/, July 16, 2018.

[27] TensorFlow, “TensorFlow,” https://www.tensorflow.org, July 16, 2018.

[28] tflearn, “tflearn,” https://www.github.com/tflearn/tflearn, July 16, 2018.

[29] SocialEngine, “SocialEngine,” https://www.socialengine.com, July 16, 2018.

[30] Google Hangouts, “Hangouts,” https://hangouts.google.com, July 16, 2018.