Word Embeddings of Monosemous Words in Dictionary for Word Sense Disambiguation

Minoru Sasaki
Dept. of Computer and Information Sciences
Faculty of Engineering, Ibaraki University
4-12-1, Nakanarusawa, Hitachi, Ibaraki, 316-8511, Japan
Email: minoru.sasaki.01@vc.ibaraki.ac.jp

***Abstract*—In the recent past, word embedding techniques have shown to capture semantic and syntactic information of natural language which could be exploited to solve the Word Sense Disambiguation (WSD) task. Word embeddings are generated using words appearing in context. However, some co-occurrence words in context have multiple meanings and are ambiguous. Therefore, it is sometimes difficult to identify the meaning of a target word by using word embeddings of context words. In this paper, we propose to use word embeddings of monosemous words for the WSD task. We consider that word embeddings of monosemous words can contribute to determining the correct sense of a target word. Also, by using word dependency in a sentence, it is possible to capture the semantic relationship between the target word and the co-occurrence word as a feature. To evaluate the efficiency of the proposed WSD method, we show that it is effective for the WSD task to use both monosemous word information and dependency relation to the target word.**

***Keywords*-word sense disambiguation; monosemous words; word embeddings.**

I. Introduction

Typically, many words have multiple meanings, depending on the context in which they are used. Identifying the sense of a polysemous word within a given context is a fundamental problem in natural language processing. For example, an English word "bank" have different senses, such as "a commercial bank" or "a land along the edge of a river" etc. Word Sense Disambiguation (WSD) is the task of deciding the appropriate meaning of a target ambiguous word in its context [9].

To solve the computational WSD problem, it is usually formulated as a classification task, where the possible word senses are the classes. In the supervised learning method, bag-of-words features extracted from a wide context window around the target word are used. In the recent past, word embedding techniques (e.g., word2vec) have shown to capture semantic and syntactic information of natural language and improve performance of the WSD task [7].

In word2vec, word embeddings are generated using words appearing in context. However, some co-occurrence words in context have multiple meanings and are ambiguous. Therefore, it is sometimes difficult to identify the meaning of a target word by using word embeddings of context words. For example, if the polysemous word "flow" appears in the context, it is not possible to distinguish the meaning of the target word "bank". However, if the monosemous word "financial" appears in the context, it is easy to distinguish the meaning of the "bank". For the word "flow", word2vec creates a word embedding containing these multiple meanings. So, these features are not effective to distinguish a target word due to its association with polysemous words. Therefore, we would like to focus on solving this issue and explore the effective features for training WSD classifiers.

In this paper, we propose a new method for WSD using word embeddings of the monosemous words in context and word dependency. We consider that word embeddings of monosemous words can contribute to determining the correct sense of a target word. Also, by using word dependency in a sentence, it is possible to capture the semantic relationship between the target word and the co-occurrence word as a feature. We show that word embeddings of monosemous words in dependency relation to the target word is effective for word sense disambiguation.

The rest of this paper is organized as follows. Section I I is devoted to the related work in the literature. Section III describes the proposed WSD methods using word embeddings of the monosemous words. In Section IV, we describe an outline of experiments and experimental results. Finally, Section V concludes the paper.

II. Related Works

Numerous works have recently demonstrated the effectiveness of bag-of-words model on WSD tasks. In supervised WSD, each occurrence of a polysemous word is converted into a small number of local features that include co-occurrence and part-of-speech information near the target word [14]. In this paper, we focus on supervised WSD using word embeddings.

Word embeddings are low-dimensional vector representations of words, based on the distributional contexts in which words appear. Word embeddings are effective at capturing intuitive characteristics of the words and can be generally useful in many NLP tasks [4][11]. Word embeddings as local context features have been used in supervised learning approaches [13].

Monosemous words can be employed to represent word contexts. Li et al. proposed the Chinese WSD method using monosemous words as features [6]. However, this method can only use limited monosemous words obtained from the Chinese thesaurus Cilin and does not use word embeddings based on neural networks. Moreover, the effectiveness of monosemous words was not verified in the Japanese WSD task. Li et al. point out that the WSD system tends to have low precision when the usage of a polysemous word is inconsistent with the monosemous words in the same class.

To obtain precise usage information, syntactic information, such as dependency relations of words has been employed. Some works exploited the dependency relations represented by the linguistic unit called bunsetsu [5][8]. These researches report that the syntactic relations are effective for WSD and document retrieval tasks. In our WSD method, we employ word embeddings of the monosemous words in context and word dependency as features and evaluate the efficiency of this WSD method.

III. WSD Methods

A. Task Description

A WSD system is used to select the appropriate sense for a target polysemous word in context. WSD can be viewed as a classification task in which each target word should be classified into one of the predefined existing senses. Word senses were annotated in a corpus in accordance with "Iwanami's Japanese Dictionary (The Iwanami Kokugo Jiten)". It has three levels for sense IDs and the middle-level sense is used in this task.

In this paper, supervised classification is employed for this WSD task. This supervised method requires a corpus of manually labeled training data to construct classifiers for every polysemous word. Then, each obtained classifier is applied to a set of unlabeled examples.

B. Supervised WSD methods

In this section, we briefly describe the baseline WSD method and our three WSD method using word embeddings of the monosemous words in context and word dependency. The first method is the WSD method using word embeddings of the only monosemous words in context. The second one is the WSD method using word embeddings of the words that have direct dependency relations with the target word. The third one is the WSD method using word embeddings of both the monosemous words and the words that have direct dependency relations with the target word.

In our experiments, we use the supervised learning approach to obtain the WSD models. The training set used to learn the models contains a set of examples in which a given target word is manually tagged with a sense. Each sentence is segmented into words by a morphological analyzer. Part-of-speech tags are assigned to the obtained words that are lemmatized.

1) The Baseline System: The baseline system uses word embeddings of the words in a sentence. In this baseline system, we calculate the average of word embeddings of all words except the target word in a sentence. Then, a supervised WSD classifier for the target word is constructed from a training set of the average vectors of input sentences and their appropriate sense label (Figure 1).

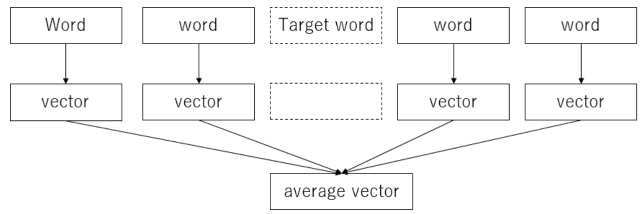

Fig. 1. Baseline System.

2) WSD using word embeddings of the only monosemous words: This WSD system employs word embeddings of the only monosemous words in context. A monosemous word is defined as as a word that has only one meaning in the "Iwanami's Japanese Dictionary (The Iwanami Kokugo Jiten)". In this system, we extract monosemous words in the two words either side of the target word and represent their word embeddings. Then, a WSD classifier for the target word is constructed from a training set of their word embeddings and their appropriate sense label (Figure 2).

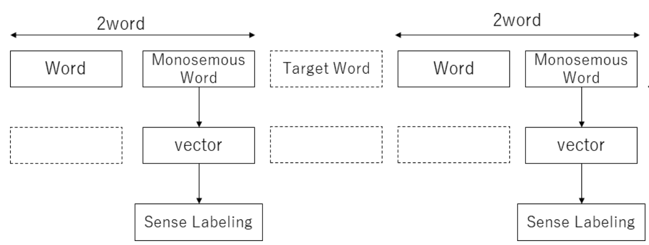

Fig. 2. WSD Using Word Embeddings of the Only Monosemous Words.

3) WSD using dependency relations with the target word: In this WSD system, we employ word embeddings of the words that have direct dependency relations with the target word. We extract co-occurrence words that have dependency relations with the target word and represent their word embeddings. We calculate the average of word embeddings of their co-occurrence words. Then, a WSD classifier for the target word is constructed from a training set of the average vectors of input sentences and their appropriate sense label (Figure 3).

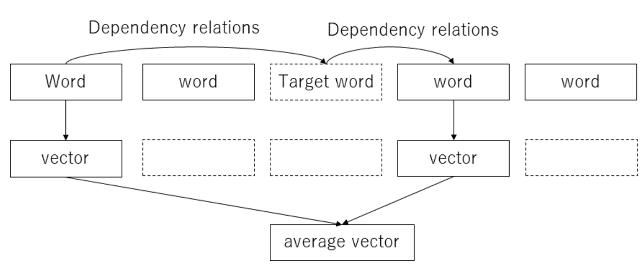

Fig. 3. WSD Using Dependency Relations with the Target word.

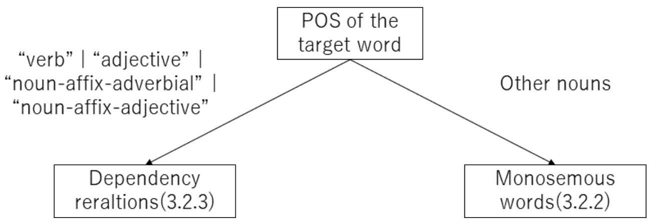

Fig. 4. WSD Using Both of Two Methods.

TABLE I
EXPERIENTIALE RESULTS OF APPLYING THE FOUR METHODS

Methods	Ave. Precision
Baseline (3.3.1)	70.16%
Monosemous (3.3.2)	68.40%
Dependency (3.3.3)	70.56%
Mono+Dep (3.3.4)	**72.08%**

4) WSD using both of the above two methods: In this WSD method, we use both of the above two methods. According to the part-of-speech of the target word, we select which method to use from the above methods. If the part of speech of the target word is "verb", "adjective", "noun-affix-adverbial" or "noun-affix-adjective", we use the WSD method that mentioned in the Section III-B2. If the part of speech of the target word is the other nouns, we use the WSD method that mentioned in the Section III-B3 (Figure 4).

IV. Experiments

To evaluate the efficiency of the proposed WSD method using word embeddings of the monosemous words in context and word dependency, we conduct some experiments to compare with the result of the baseline system. In this section, we describe an outline of the experiments.

A. Data Set

We use the Semeval-2010 Japanese WSD task data set, which includes 50 target words comprising 22 nouns, 23 verbs, and 5 adjectives [10]. In this data set, there are 50 training and 50 test instances for each target word.

B. Word Vector Representations

In these experiments, we use the two available pre-trained Japanese word embeddings. The first set of word vectors is "nwjc2vec" [12]. The nwjc2vec is pre-trained word embeddings constructed from NINJAL Web Japanese Corpus using word2vec. The second set is "Asahi Shimbun Word Vectors"[1]. This set is constructed from about 8 millions newspaper articles from Asahi Shimbun, which is a Japanese newspaper.

C. Preprocessing

Semantic and Syntactic features are extracted from the context of the target word (two words to the right and left) as described in the previous section. Each sentence of training data and test data is segmented into words by a morphological analyzer. As a morphological analyzer, we use MeCab[3] to obtain words and their part-of-speech. To obtain dependency relations for all words in a sentence, we use Cabocha[2] as a syntactic analyzer. Moreover, to improve performance, we remove words used as noun suffix and affix, and Japanese stop words from context words, such as "こと (thing)" and "様 (like)", etc.

For the obtained feature set of training data, we construct classification model using Support Vector Machine (SVM). When the classification model is obtained, we predict one sense for each test example using this model. To employ the SVM for distinguishing more than two senses, we use one-versus-rest binary classication approach for each sense. As a result of the classification, we obtain precision value of each method to analyze the average performance of systems.

TABLE II
EXPERIENTIAL RESULTS OF APPLYING THE THREE TYPES OF WORD EMBEDDINGS.

Vectors	Baseline	Mono+Dep(3.3.4)
asahi(skip-gram)	69.52%	**70.04%**
asahi(cbow)	69.16%	**69.96%**
asahi(glove)	69.20%	**70.60%**
nwjc2vec	70.16%	**72.08%**

D. Experimental Results

Table I shows the results of the experiments of applying the four methods in the previous section. According this table, the proposed methods using word embeddings of the only monosemous words and using dependency relations with the target word achieve better results than the baseline system. However, the WSD method using word embeddings of the only monosemous words does not achieve improvement over the baseline system. As the results of these experiments, word embeddings of the monosemous words are effective for noun word sense disambiguation task except for noun-common-adverb and noun-adjective-base form. If the target word is verb, adjective and noun (noun-common-adverb and noun-adjective-base form), word embedding features of co-occurrence words are not so effective to capture the characteristics of context.

Regardless of the part-of-speech of the target word, word embeddings of the words that have direct dependency relations with the target word are effective to obtain context information. In this way, by selecting the WSD method according to the part-of-speech of the target word, we consider that the average precision of the all target words can be increased.

Moreover, we now show that the proposed method can be applied to other word embeddings. To do so, we use a word embedding based on the "Asahi Shimbun Word Vectors". Table II shows the results of the experiments of applying the three types of word embeddings in the "Asahi Shimbun Word Vectors" (skip-gram, CBOW and GloVe). The proposed methods using these word embeddings achieve better results than the baseline system. However, the average precision of the WSD system slightly decreased in compared with the method using nwjc2vec.

V. Conclusion and Future Work

In this paper, we proposed a new method for WSD using word embeddings of the monosemous words in context and word dependency. The efficiency of the proposed method was evaluated on the Semeval-2010 Japanese WSD task dataset. The results showed that the proposed methods using word embeddings of the only monosemous words and using dependency relations with the target word achieve better results than the baseline system.

In the future, we will analyze the dependency relation and the co-occurrence relation between monosemous words and polysemous words to investigate the effectiveness of monosemous words for word sense disambiguation. Moreover, for providing more useful sense information, we will construct a lexical semantic resource which is useful for expressing the target relation of monosemous words.

References

[1] "Asahi shimbun word vectors," http://www.asahi.com/shimbun/medialab/word_embedding/, [accessed October 2018].

[2] "Cabocha: Yet another japanese dependency structure analyzer," http://taku910.github.io/cabocha/, [accessed October 2018].

[3] "Mecab: Yet another part-of-speech and morphological analyzer," http://taku910.github.io/mecab/, [accessed October 2018].

[4] X. Chen, Z. Liu, and M. Sun, "A unified model for word sense representation and disambiguation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, 2014, pp. 1025–1035.

[5] P. Kathuria and K. Shirai, "Word sense disambiguation based on example sentences in dictionary and automatically acquired from parallel corpus," in *Proceedings of the Advances in Natural Language Processing: 8th International Conference on NLP (JapTAL2012)*, H. Isahara and K. Kanzaki, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 210–221.

[6] J. Li and C. Huang, "A model for word sense disambiguation," in *International Journal of Computational Linguistics & Chinese Language Processing, Volume 4, Number 2, August 1999*, 1999, pp. 1–20.

[7] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, ser. NIPS'13. USA: Curran Associates Inc., 2013, pp. 3111–3119.

[8] K. Nagamatsu and H. Tanaka, "Evaluation of a similarity measure based on co-occurrence and dependency between words," in *Technical report of the Special Interest Group on Natural Language Processing of the Information Processing Society in Japan (IPSJ-NL) (in Japanese)*, vol. 116-11, 1996.

[9] R. Navigli, "Word sense disambiguation: A survey," *ACM Computing Surveys*, vol. 41, no. 2, pp. 10:1–10:69, Feb. 2009.

[10] M. Okumura, K. Shirai, K. Komiya, and H. Yokono, "Semeval-2010 task: Japanese wsd," in *Proceedings of the 5th International Workshop on Semantic Evaluation*, ser. SemEval '10. Stroudsburg, PA, USA: Association for Computational Linguistics, 2010, pp. 69–74.

[11] M. Sasaki, K. Komiya, and H. Shinnou, "Efficiency of the dictionary definition for word sense disambiguation based on word embeddings," in *22nd Annual Meeting of the Association of Natural Language Processing (in Japanese)*, 2016, pp. 449–452.

[12] H. Shinnou, M. Asahara, K. Komiya, and M. Sasaki, "nwjc2vec: Word embedding data constructed from ninjal web japanese corpus," in *Journal of Natural Language Processing*, vol. 24-5, 2017, pp. 705–720.

[13] H. Sugawara, H. Takamura, R. Sasano, and M. Okumura, "Context representation with word embeddings for wsd," in *Computational Linguistics: 14th International Conference of the Pacific Association for Computational Linguistics, PACLING 2015*. Singapore: Springer Singapore, 2015, pp. 108–119.

[14] D. Yarowsky, "Unsupervised word sense disambiguation rivaling supervised methods," in *Proceedings of the 33rd Annual Meeting on Association for Computational Linguistics*, ser. ACL '95. Stroudsburg, PA, USA: Association for Computational Linguistics, 1995, pp. 189–196.