Japanese Word Sense Disambiguation Using Gloss Information of a Japanese Dictionary

Hiroki Okemoto
Major in Computer and Information Sciences
Graduate School of Science and Engineering,
Ibaraki University
e-mail: 21nm716g@vc.ibaraki.ac.jp
4-12-1, Nakanarusawa, Hitachi, Ibaraki, Japan

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

Abstract— **Word Sense Disambiguation (WSD) aims to find an appropriate meaning of ambiguous words in a particular context. Traditional supervised WSD methods rarely take into account lexical resources, such as WordNet, but recent studies have shown the effectiveness of incorporating glosses into neural networks for WSD. However, since most of this research is based on WSD, such as English, it has not been shown whether Japanese gloss information, such as a Japanese dictionary is effective for WSD. In this study, we aim to evaluate the effectiveness of using glossary information of the Japanese dictionary for WSD. As results of experiments, we found it effective to use glosses of the Japanese dictionary in WSD.**

Keywords-WSD; Japanese Dictionary; Machine Learning.

I. INTRODUCTION

In this section, we present the purpose and background of our research.

A. *Research background*

A word may have multiple meanings depending on its context. For example, the word "合う" has multiple meanings, such as "同じ動作をする" (Do the same action), such as "あなたと話し合う"(Talk to you) and "一致する"(Match), such as "意見が合う"(To agree in opinion). Thus, there is WSD, a basic task of Natural Language Processing (NLP) aimed at finding the exact meaning of ambiguous words in a particular context.

WSD has been studied in various ways[1][2] to date, and there are several approaches. Knowledge-based techniques [3] use lexical knowledge, such as glosses, to infer the correct meaning of the meaning of ambiguous words in context. However, the biggest drawback of knowledge-based methods is that they perform worse than supervised methods. Also, supervised methods usually train separate classifiers for words. Therefore, we cannot easily extend to the WSD task of words which ambiguate all polysemes in the text. In addition, in the neural-based method, only the local context of the target word is considered, and it becomes impossible to distinguish the minority meaning which is not in the training data.

In recent years, Huang et al. [4] conducted experiments using several English word WSD benchmark datasets with glosses using a technique called “GlossBERT” to construct context-gloss pairs and showed that this approach significantly outperformed state-of-the-art systems.

B. *Purpose of research*

In this study, in order to solve the problem of the above background, the purpose is to analyze whether the use of the gloss of the Japanese dictionary is effective or not by creating data in which the gloss of the Japanese dictionary is combined with the example sentence whose semantic meaning is known in the WSD system. By using glosses in the Japanese dictionary, it is possible to capture information about meaning that does not exist in the training data, and it is possible to capture detailed meaning differences between meanings.

In Section 2, we describe the proposed method. In Section 3, we describe the experiments conducted in this research. In Section 4, we discuss the experimental results in Section 3. In Section 5, we describe the conclusion and future work of this research.

II. METHODS

This section describes the proposed WSD using glosses of Japanese dictionaries in WSD systems.

A. *System overview*

The WSD method proposed in this paper uses glosses of Japanese dictionaries in WSD systems. We show the general execution order of the proposed method in Figure 1.

First, training data combining Japanese dictionary glosses and example sentences are created.

Next, the prepared training data and the test data to be compared are converted into a context vector by using BERT (NWJC-BERT). Therefore, the data was morphemically analyzed and converted to a lexeme of Unidic, and then converted to a context vector by BERT.

Then, the cosine similarity between the CLS vector of the converted training data and the object word vector of the test data, and the cosine similarity between the object word vector of the training data and the object word vector of the test data are respectively obtained, and their average values are obtained, and the values become evaluation values. The number of evaluation values obtained in this process is the number of training data and the number of combinations of test data per target word.

Finally, we compare the obtained evaluation values, and the meaning of the target word of the training data corresponding to the maximum value becomes the meaning of the target word of the test data.

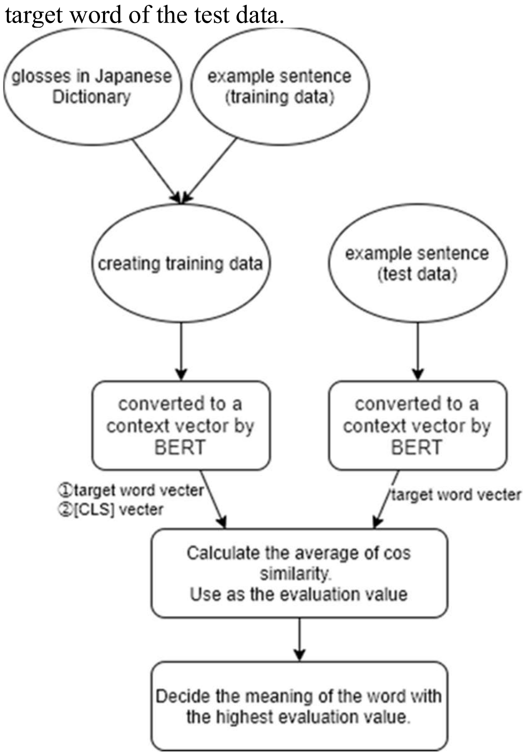


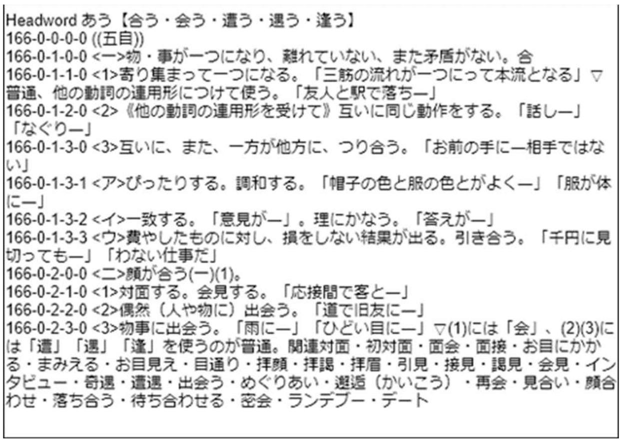
Figure 1. The general execution order of the proposed method

B. Description of the usage data

The target words in this study are 50 target words that are SemEval 2010 Japanese WSD task data (Okumura, Shirai, Komiya, Yokono, 2010). As the use data, 50 pieces of example sentence data using the word as training data and test data are respectively prepared from the modern Japanese written language balanced corpus (BCCWJ), and the Iwanami Japanese dictionary is used as the Japanese dictionary. This dictionary is the data distributed in SemEval -2010: Japanese WSD task.

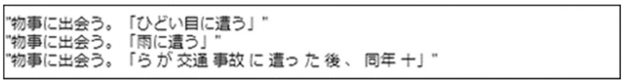
C. Preparation of training data

First, when creating training data using the meaning glosses of the Japanese dictionary this time, the meaning and example sentence are extracted from the meaning definition sentence of the Iwanami Japanese dictionary. The format of the Iwanami Japanese Dictionary used this time is shown in Figure 2. The part enclosed by "「」" in the gloss is an example sentence, and the "-" part is used as one example sentence by complementing with a headword. In addition, the parts enclosed by "(())", "<<>>" and the parts following "▽" were judged to be irrelevant and removed. A sentence separated by the rest of the punctuation marks is used as one meaning. For example, if it is 166 − 0 -2 -3 − 0 in Figure 2, the meaning of the word is "物事に出会う。"(Encounter things), and the example sentence is "雨に遭う"(Get rain) or "ひどい目に合う"(Go through a bitter experience). Next, the definitions extracted from the Japanese dictionary and example sentences of example sentences and training data extracted from the Japanese dictionary are combined one by one to form one data. The format of the data is "gloss「example sentence」" as shown in Figure 3.



Headword あう【合う・会う・遭う・遇う・逢う】
166-0-0-0-0 ((五自))
166-0-1-0-0 <一>物・事が一つになり、離れていない、また矛盾がない。合
166-0-1-1-0 <1>寄り集まって一つになる。「三筋の流れが一つにって本流となる」▽普通、他の動詞の連用形につけて使う。「友人と駅で落ち—」
166-0-1-2-0 <2>《他の動詞の連用形を受けて》互いに同じ動作をする。「話し—」「なぐり—」
166-0-1-3-0 <3>互いに、また、一方が他方に、つり合う。「お前の手に—相手ではない」
166-0-1-3-1 <ア>ぴったりする。調和する。「帽子の色と服の色とがよく—」「服が体に—」
166-0-1-3-2 <イ>一致する。「意見が—」。理にかなう。「答えが—」
166-0-1-3-3 <ウ>費やしたものに対し、損をしない結果が出る。引き合う。「千円に見切っても—」「わない仕事だ」
166-0-2-0-0 <二>顔が合う(一)(1)。
166-0-2-1-0 <1>対面する。会見する。「応接間で客と—」
166-0-2-2-0 <2>偶然（人や物に）出会う。「道で旧友に—」
166-0-2-3-0 <3>物事に出会う。「雨に—」「ひどい目に—」▽(1)には「会」、(2)(3)には「遭」「遇」「逢」を使うのが普通。関連対面・初対面・面会・面接・お目にかかる・まみえる・お目見え・目通り・拝顔・拝謁・拝眉・引見・接見・謁見・会見・インタビュー・奇遇・遭遇・出会う・めぐりあい・邂逅（かいこう）・再会・見合い・顔合わせ・落ち合う・待ち合わせる・密会・ランデブー・デート

Figure 2. Format of Iwanami Japanese Dictionary



"物事に出会う。「ひどい目に遭う」"
"物事に出会う。「雨に遭う」"
"物事に出会う。「ら が 交通 事故 に 遭っ た 後、 同年 十」"

Figure 3. Format of the created training data

D. Cosine similarity

In this study, we use cosine similarity [5] as a method to calculate similarity between vectors. We can calculate the cosine similarity with the following equation. The closer the maximum value is to 1, the more similar the vectors are.

$$\cos(\vec{p}, \vec{q}) = \frac{\vec{p} \cdot \vec{q}}{|\vec{p}| \cdot |\vec{q}|}$$

In the present method, a plurality of vectors, such as a CLS vector of training data and a target word vector of test data, and a target word vector of training data and a target word vector of test data are compared by one combination of the training data and the test data. Therefore, we determined the cosine similarity and used the average value as the evaluation value.

III. EXPERIMENT

In this section, we present the objectives, methods, and results of this experiment.

A. Purpose of the experiment

In this study, glosses of 50 test data sentences of 50 target words (Okumura, Shirai, Komiya, Yokono, 2010), which are

SemEval 2010 Japanese WSD task data, are determined up to the middle classification of Iwanami Japanese Dictionary using glosses of Japanese dictionaries in WSD systems. By doing so, we aim to verify the effective-ness of using glosses in Japanese dictionaries compared to when they are not used.

B. *Experimental Methods*

An experimental method based on the proposed method is presented.

1) Creating data using glosses

In this experiment, we experimented by changing the method of separating the example sentences of the training data, which was added after the meaning. This is because the length of the example sentence in the training data is not constant, and it is expected that when the sentence becomes long, there will be parts which are not related to the judgment of meaning. Table 1 below shows how to separate the example sentences used this time.

TABLE I. LIST OF HOW TO SEPARATE EXAMPLE SENTENCES

case	how to separate example sentences
A	not separate
B	separate by punctuation. And the part that contains the target word.
C	A total of 7 words, including 3 words before and after the target word
D	A total of 11words, including 5 words before and after the target word
E	A total of 15words, including 7 words before and after the target word
F	A total of 19 words, including 9 words before and after the target word

2) Target for comparison with test data

In this experiment, the average value of the cosine similarity between the target word vector of the test data and the target of comparison of the training data is used as the evaluation value. However, we are experimenting with different targets for comparison. In this study, we conducted experiments using two patterns: one with CLS vector and the other without CLS vector. This is because, since the target word is basically in the example sentence, the target word vector is considered to be more influenced by words around the example sentence, and the CLS vector is considered to reflect more the semantic context when it is added to the vector comparison object of the whole sentence.

3) Evaluation Method

In this experiment, the test data consists of 50 words as de-scribed in Section 2.2, and there are 50 data items per word. We compared them with the training data, determined the evaluation value, determined the meaning of the training data which became the maximum value as the meaning of the test data, and obtained the correct answer rate. We then determined the average of the correct answers of 50 words.

C. *Results*

First, Table 2 below shows the average of the correct answer rates of 50 words using training data that does not use Japanese dictionary glosses, which is the object of comparison with this method.

TABLE II. AVERAGE OF CORRECT ANSWER RATE OF WSD WITHOUT GLOSSES OF JAPANESE DICTIONARY

case	[CLS] vec + target word vec	target word vec
A	78.68%	79.00%

In order to show the effectiveness of glosses, we compare it with the method of supervised learning which does not use glosses of Japanese dictionary. We use a Multi-Layer Perceptron (MLP) as a classification model to learn training data and estimate the glosses of test data. In this method, the number of nodes in the middle layer is set to 50, the stochastic gradient descent method is used as an optimization method, the number of epochs as the number of learning iterations is set to 50, and the batch size is set to 200. Since the training data was small and the values were unstable with each execution, we conducted six tests to find the average value. The experimental results are shown in Table 3 below.

TABLE III. SUPERVISED LEARNING WITHOUT GLOSSES（MLP）

	correct answer rate
1	68%
2	63%
3	58%
4	64%
5	68%
6	61%
Ave	64%

Next, the experimental results based on this experimental method are shown in Table 4 below.

TABLE IV. AVERAGE OF THE CORRECT ANSWER RATES OF WSD USING GLOSSES OF JAPANESE DICTIONARIES

case	[CLS] vec + target word vec	target word vec
A	78.88%	78.32%
B	79.28%	79.16%
C	78.36%	79.24%
D	78.52%	79.16%
E	79.20%	79.68%
F	78.28%	79.44%

As a result, the highest percentage of correct answers was 79.28% for "case B" in the case of “CLS vector and target word vector” and 79.68% for “target word vector” in the case of “case E”.

Comparing Table 2, Table 3, and Table 4, it can be seen that the method using glosses is more accurate.

IV. DISCUSSION

From the experimental results, the WSD using the gloss of the Japanese dictionary was slightly more correct than the WSD without the gloss. From this, we believe that it is effective to use glosses in WSD.

When the division method was changed, the word with the highest correct answer rate was "case B" with 79.28% in the case of "CLS vector and target word vector," and the word with the highest correct answer rate was "case E" with 79.68% in the case of "target word vector.". We think that this is because not only the gloss but also the example sentence had to be established as a sentence to some extent in order to consider the CLS vector and the target CLS word vector. In addition, in the case of the “target word vector”, since a higher rate of correct answers is generally obtained when the words are separated by the number of words before and after the punctuation mark than when they are separated by the punctuation mark, we think that it is effective for example sentences to always have the context of words.

Based on these results, we think that by punctuating with the number of preceding and following phrases, we can maintain the context of the word to some extent while making it possible to maintain the context of the word, and thus we can further increase the rate of correct answers.

This time, the meaning of the word was judged by obtaining the cosine similarity of the object word vector. However, we expect improvements by using of MLP for semantic analysis. In addition, there was a bias in the number of examples depending on the meaning of the word. In addition, we expect improvements by increasing the amount of data and using related words because there are few or short definitions for a single meaning.

V. CONCLUSION AND FUTURE WORK

In this study, we conducted experiments to analyze whether it is effective to use glosses of Japanese dictionaries in WSD systems. In the experiment, we divided the comparison object with the object word vector of the test data into two types, the case in which the object word vector of the training data and CLS vector of the training data are included and the case in which it is not included. In addition, when combining meaning and example sentences, we changed the way of separating example sentences into six types.

As a result of the experiment, it was possible to obtain a higher correct answer rate of the data using the gloss, when the object to be compared with the object word vector of the test data includes the object word vector of the training data and the CLS vector of the training data, and when it does not include them. Therefore, we confirmed the effectiveness of using glosses of Japanese dictionaries in WSD systems.

Future work will include using other semantic methods, such as MLP, scrutinizing the data used, and increasing the amount of data.

REFERENCES

[1] Fuli Luo, Tianyu Liu, Qiaolin Xia, Baobao Chang, Zhifang Sui, “Incorporating Glosses into Neural Word Sense Disambiguation”, Association for Computational Linguistics

[2] Fuli Luo , Tianyu Liu , Zexue He , Qiaolin Xia , Zhifang Sui , Baobao Chang, “Leveraging Gloss Knowledge in Neural Word Sense Disambiguation by Hierarchical Co-Attention”, Association for Computational Linguistics

[3] Andrea Moro, Alessandro Raganato, Roberto Navigli ,“Entity Linking meets Word Sense Disambiguation: a Unified Approach”, Transactions of the Association for Computational Linguistics

[4] Luyao Huang, Chi Sun, Xipeng Qiu, Xuanjing Huang,”GlossBERT: BERT for Word Sense Disambiguation with Gloss Knowledge”, Association for Computational Linguistics

[5] Amit Singhal,”Modern Information Retrieval: A Brief Overview”, Bulletin of the IEEE Computer Society Technical Committee on Data Engineering