Japanese Word Sense Disambiguation Using Paraphrasing and Machine Learning

Yuma Toda, Masaki Murata
Department of Information and Electronics
Tottori University
4-101 Koyama-Minami, Tottori 680-8552, Japan
Email: {s122033,murata}@ike.tottori-u.ac.jp

Qing Ma
Department of Applied Mathematics and Informatics
Ryukoku University
Seta, Otsu, Shiga 520-2194, Japan
Email: qma@math.ryukoku.ac.jp

***Abstract*—In this study, we performed word sense disambiguation using machine learning. We increased the number of data items of a training dataset by semi-automatically constructing new training data items using paraphrases of an ambiguous word. We performed word sense disambiguation using machine learning with the constructed training dataset. Through our experiments, we confirmed that relatively higher accuracies can be obtained using the additional constructed training dataset. A support vector machine using the constructed training dataset obtained the highest accuracy (0.78). Furthermore, we performed analysis of features used in machine learning and clarified useful features for estimating senses.**

I. Introduction

Word sense disambiguation is an important problem in natural language processing. It involves the selection of an appropriate sense among multiple senses of an ambiguous word. Word sense disambiguation is useful for machine translation and knowledge extraction. According to a study by Shin'no et al. [1], approximately 70% of errors in word sense disambiguation are because of the lack of data items in training data.

In this study, we increase the number of data items in a training dataset by semi-automatically constructing data items of a training dataset using paraphrases of an ambiguous word. We perform word sense disambiguation using machine learning with the constructed training dataset. We perform a study on Japanese words.

The characteristics of this study are as follows.

- We increased the number of data items in a training dataset using sentences containing manually selected paraphrases.
- Generally, in machine learning, the distribution of the data items among categories is important. We decreased the numbers of the data items for categories in an additional constructed training dataset so as that the distribution at the numbers of data items for categories in an original training dataset was equal to that in the additional constructed training dataset. Our experiments confirmed that this method of changing the number of data items in an additional constructed training dataset was important.
- When changing the number of data items in an additional constructed training dataset, we confirmed through experiments that we can obtain higher accuracies using this dataset.
- We performed an analysis of features used in machine learning and clarified useful features for estimating senses. We confirmed that when using an additional constructed training dataset, features can work more effectively.

II. Our proposed method

A. Word sense disambiguation

In this study, we perform word sense disambiguation using paraphrasing and machine learning, similar to the method proposed in a previous study [2]. An input word sense disambiguation is a sentence containing an ambiguous word. An output is the appropriate sense among multiple senses. We use supervised machine learning based on a training dataset, as reported in a previous study [3]. When the quantity of training data is small, word sense disambiguation is incorrectly performed. Therefore, in this study, we increase the data items of a training dataset by semi-automatically constructing such data items using paraphrases of ambiguous words. To increase the data items of a training dataset, we extract sentences containing a word similar to an ambiguous word. By transforming the similar word to the ambiguous word in extracted sentences, we can increase data items of a training dataset. We use 48 features in learning; the features contain sentence structures and words in sentences. We use the maximum entropy method [4] and a support vector machine [5], [6] for machine learning.

B. Increasing data items in a training dataset using paraphrasing

We explain the method of increasing the data items in a training dataset using paraphrasing. Suppose that X is an ambiguous word with three senses. First, we manually select a word that roughly indicates each sense in X. When we select a word as in the above procedure, we consult the words used in the definition sentence of a sense. Although we often select a word among words in the definition sentence, we sometimes select a word that is not in the definition sentence, and we manually associate it with the definition sentence. Suppose that $x1$, $x2$, and $x3$ are words that roughly indicate the first, second, and third sense. We extract sentences containing $x1$,

TABLE I: Sentences before and after we transform a word to another word

Word	Sentences containing *naiyou*, *douki*, or *kachi*	Sentences after we transformed the word to *imi* (meaning)	Sense
naiyou (content)	*Naiyou-wa bekkou-no toori-dearu.* (The contents are as it is a separate section.)	*Imi-wa bekkou-no toori-dearu.* (The meanings are as it is a separate section.)	Sense 1
douki (motive)	*Chakuriku-no douki-wa akirakani-sareteinai.* (The motive of the landing has not been clarified.)	*Chakuriku-no imi-wa akirakani-sareteinai.* (The meaning of the landing has not been clarified.)	Sense 2
kachi (value)	*Ippyou-no kachi-ga mottomo hikui.* (The value of one vote is the lowest.)	*Ippyou-no imi-ga mottomo hikui.* (The meaning of one vote is the lowest.)	Sense 3

$x2$, or $x3$ from a text corpus (i.e.,., newspaper articles). We transform $x1$, $x2$, and $x3$ to X in the extracted sentences. When $x1$ is transformed into X, the transformed X is an X with the first sense. The transformed sentence is used as a data item of a training dataset. Using this procedure, we can increase the number of training data items. We perform word sense disambiguation of a word using the additional constructed training dataset.

This study differs from the previous study [2] in the language studied. The previous study [2] dealt with English, and this study deals with Japanese. In addition, in this study, there are cases in which we manually select a word that is not in a definition sentence and associate it with a definition sentence.[1]

An example of increasing data items of a training dataset using paraphrasing is as follows.

We increase data items of a training dataset for the ambiguous word *imi* (meaning). This word has the following three senses in the Iwanami Kokugo Japanese dictionary.

Sense 1:
Sono kotoba-no arawasu naiyou. Igi.
(the word) (indicate) (content) (sense or significance)
(The contents the word indicates. Sense.)

Sense 2:
Hyougen-ya koui-no ito, douki.
(expression) (action) (aim) (motive)
(Aim and motive of expression or action.)

Sense 3:
Hyougen-ya koui-no motsu kachi. Igi.
(expression) (action) (have) (value) (sense or significance)
(The value that the expression and act have. The significance.)

We manually select words that roughly indicate three senses. We select *naiyou* (content), *dougi* (motive), and *kachi* (value) as words roughly indicating the three senses. Then, we extract sentences containing *naiyou*, *dougi*, and *kachi* from a text corpus. Examples of extracted sentences are as follows:

Example sentences including *naiyou*:
Naiyou-wa bekkou-no toori-dearu.
(content) (separate section) (be as it is)
(The contents are as it is a separate section.)

Example sentences including *douki*:
Chakuriku-no douki-wa akirakani-sareteinai.
(landing) (motive) (has not been clarified)
(The motive of the landing has not been clarified.)

Example sentences including *kachi*:
Ippyou-no kachi-ga mottomo hikui.
(one vote) (value) (the most) (low)
(The value of one vote is the lowest.)

When we transform *naiyou* to *imi*, the transformed *imi* is intended to be *imi* with Sense 1. Because the transformed sentence can be used as a data item of a training dataset, we can increase the number of data items in a training dataset. We show some examples of constructing training data items and increasing the number of data items in a training dataset by transforming *naiyou*,*douki*, and *kachi* into *imi* in Table I. We perform word sense disambiguation of the word *imi* using the additional constructed training dataset.

C. Maximum entropy method

In our study, we use maximum entropy method and support vector machine as machine learning. In this section, we explain maximum entropy method. In the next section, we explain support vector machine.

In the ME method [4], the distribution of probabilities $p(a,b)$ is calculated for the case where Equation (1) is satisfied and Equation (2) is maximized; the desired probabilities $p(a|b)$ are then calculated using the distribution of probabilities $p(a,b)$:

$$\sum_{a\in A,b\in B} p(a,b)g_j(a,b) = \sum_{a\in A,b\in B} \tilde{p}(a,b)g_j(a,b) \quad for\ \forall f_j\ (1 \le j \le k) \tag{1}$$

$$H(p) = -\sum_{a\in A,b\in B} p(a,b)\ log\,(p(a,b)), \tag{2}$$

where A, B, and F are, respectively, sets of categories, contexts, and features $f_j(\in F, 1 \le j \le k)$; $g_j(a,b)$ is a function defined as 1 when context b has feature f_j and the category is a, or defined as 0 otherwise; and $\tilde{p}(a,b)$ is the occurrence rate of (a,b) in the training dataset.

D. Support vector machine method

In this method, data comprising two categories is classified by dividing the space with a hyperplane. When the margin between examples that belong to one category and examples that belong to the other category in the training dataset is larger, the probability of incorrectly choosing categories in the open data is smaller. The hyperplane that maximizes the margin is determined and classification is performed using this hyperplane.

[1]Examples in which we manually select a word that is not in a definition sentence and associate it with a definition sentence are *yoso* (foreign) and *shika* (only) described in Section III-B.

The SVM method used in this study is implemented by combining the SVM method [6] and the pair-wise method. We used a linear kernel function in the SVM method.

III. Experiments

A. How to perform experiments

An input is a sentence containing an ambiguous word. An output is an appropriate sense among plural senses. In our experiments, we use four nouns among 50 ambiguous words in SemEval 2. SemEval 2 is a dataset prepared for a contest of word sense disambiguation [7]. SemEval 2 is constructed manually to enable its use in word sense disambiguation experiments. 50 training data items and 50 test data items are prepared for an ambiguous word.

By changing training datasets used in experiments and comparing the experimental results using different training datasets, we confirm the effectiveness of increasing the quantity of a training dataset using paraphrasing. We use three kinds of training datasets: "only a training dataset in SemEval 2," "both a training dataset in SemEval 2 and an additional training dataset based on paraphrasing," and "only an additional training dataset based on paraphrasing." In evaluation, the appropriate sense is judged to be correct.

In the cases using an additional constructed training dataset based on paraphrasing, we conduct the following two kinds of experiments:

- Not changed:
  We use an additional constructed training dataset without changing it.
- Changing the number of data items in an additional training dataset:
  Generally, in machine learning, the distribution of the data items among categories is important. We decrease the number of data items in categories so as to equalize the distribution of a training dataset for categories in SemEval 2 and that in an additional constructed training dataset. For example, suppose that the numbers of data items for Senses 1 and 2 are 30 and 20, respectively, in a training dataset of SemEval 2, and the numbers of data items for Senses 1 and 2 are 300 and 250 in an additional constructed training dataset. In this case, we decrease the number of data items for Sense 2 in the additional constructed training dataset to 200.

We use the maximum entropy method and a support vector machine for machine learning. Tables II and III show the features (the information used in machine learning) used in the experiments. The tables draw on previous papers [8], [9]. These features are extracted from a sentence containing a treated word (an ambiguous word).

The category number in the table is a ten-digit number described in the Japanese thesaurus Bunrui Goi Hyou [10], [11]. Words with similar meanings have similar ten-digit numbers. In this study, we use the first five and three digits of a number as features. Therefore, we use the upper concept of each word as features.

TABLE II: Features used in machine learning

ID	Explanation of feature
F1	Nouns in sentences
F2	The three word just before and just after the target word
F3	The first 1, 2, 3, 4, 5, and 7 digits of the category number of F2
F4	The functional words in the bunsetsu containing a target word
F5	The POS of F4
F6	The first 1, 2, 3, 4, 5, and 7 digits of the category number of F4
F7	The first functional word in the bunsetsu containing a target word
F8	The POS of F7
F9	The first 1, 2, 3, 4, 5, and 7 digits of the category number of F7
F10	The last functional word in the bunsetsu containing a target word
F11	The POS of F10
F12	The first 1, 2, 3, 4, 5, and 7 digits of the category number of F10
F13	The content words in the bunsetsu modifying the bunsetsu containing a target word
F14	The POS of F13
F15	The first 1, 2, 3, 4, 5, and 7 digits of the category number of F13
F16	The functional words in the bunsetsu modifying the bunsetsu containing a target word
F17	The POS of F16
F18	The first 1, 2, 3, 4, 5, and 7 digits of the category number of F16
F19	The first content word in the bunsetsu modifying the bunsetsu containing a target word
F20	The POS of F19
F21	The first 1, 2, 3, 4, 5, and 7 digits of the category number of F19
F22	The last content word in the bunsetsu modifying the bunsetsu containing a target word
F23	The POS of F22
F24	The first 1, 2, 3, 4, 5, and 7 digits of the category number of F22
F25	The first functional word in the bunsetsu modifying the bunsetsu containing a target word
F26	The POS of F25
F27	The first 1, 2, 3, 4, 5, and 7 digits of the category number of F25
F28	The last functional word in the bunsetsu modifying the bunsetsu containing a target word
F29	The POS of F28
F30	The first 1, 2, 3, 4, 5, and 7 digits of the category number of F28
F31	The content words in the bunsetsu modified by the bunsetsu containing a target word
F32	The POS of F31
F33	The first 1, 2, 3, 4, 5, and 7 digits of the category number of F31
F34	The functional words in the bunsetsu modified by the bunsetsu containing a target word
F35	The POS of F34
F36	The first 1, 2, 3, 4, 5, and 7 digits of the category number of F34
F37	The first content word in the bunsetsu modified by the bunsetsu containing a target word
F38	The POS of F37
F39	The first 1, 2, 3, 4, 5, and 7 digits of the category number of F37
F40	The last content word in the bunsetsu modified by the bunsetsu containing a target word
F41	The POS of F40
F42	The first 1, 2, 3, 4, 5, and 7 digits of the category number of F40
F43	The first functional word in the bunsetsu modified by the bunsetsu containing a target word
F44	The POS of F43

TABLE III: Features used in machine learning (cont.)

ID	Explanation of feature
F45	The first 1, 2, 3, 4, 5, and 7 digits of the category number of F43
F46	The last functional word in the bunsetsu modified by the bunsetsu containing a target word
F47	The POS of F46
F48	The first 1, 2, 3, 4, 5, and 7 digits of the category number of F46

TABLE IV: The number of data items of *imi* (meaning)

	Training data	Test data	Additional training data by paraphrasing
Sense 1	25	27	4403
Sense 2	8	10	370
Sense 3	17	12	1177
Unknown sense	0	1	0
Total	50	50	5950

B. Selection of characteristic words

In this study, we use the following nouns among the 50 treated words in SemEval 2: *imi* (meaning), *kodomo* (children), *hoka* (other), and *jouhou* (information). We manually select a word that roughly indicates a sense. The selected word, referred to as a characteristic word, is used for paraphrasing.

For *imi*, we use *naiyou*, *douki*, and *kachi* as Senses 1, 2, and 3, respectively, as described in Section II-B.

The selection of characteristic words for *kodomo* is as follows: *kodomo* has the following two senses.

Sense 1:
Osanai ko. Jidou.
(young) (child) (pupil)
(A young child. A pupil.)

Sense 2:
Jibun-no mouketa ko. Musuko, musume. ko.
(one) (provide) (child) (son) (daughter) (child)
(The child one provided. Son, daughter. Child.)

We use *jidou* (pupil) and *musuko* (son) as Senses 1 and 2, respectively.

The selection of characteristic words for *hoka* is as follows: *hoka* has the following two senses.

Sense 1:
Aru kijun, han'i ni fukumarenai bubun.
(standard) (extend) (be not included) (part)
(The part that is not included in some standards and extends.)

Sense 2:
Soreigai dewanai toiu kimochi de iu.
(besides that) (that it isn't) (feeling) (say)
(I say with the feeling that it isn't besides that.)

In the case of *hoka*, we could not select a characteristic word from definition sentences. We use *yoso* (foreign) and *shika* (only) as Senses 1 and 2, respectively. These words are associated by us manually by consulting definition sentences.

The selection of characteristic words for *jouhou* is as follows: *jouhou* has the following two senses.

TABLE V: The number of data items of *kodomo* (child)

	Training data	Test data	Additional training data by paraphrasing
Sense 1	26	18	997
Sense 2	24	32	783
Total	50	50	1780

TABLE VI: The number of data items of *hoka* (other)

	Training data	Test data	Additional training data by paraphrasing
Sense 1	49	50	490
Sense 2	1	0	11708
Total	50	50	12198

TABLE VII: The number of data items of *jouhou* (information)

	Training data	Test data	Additional training data by paraphrasing
Sense 1	4	8	157
Sense 2	46	42	477
Total	50	50	634

Sense 1:
Aru monogoto no jijou ni tsuite no shirase.
(matter) (circumstance) (about) (news)
(The news about circumstances of a matter.)

Sense 2:
Sorewo tooshite nannraka no chishiki ga
(that) (through) (some) (knowledge)
erareru younamono.
(something that we can acquire)
(Something of which we can acquire some knowledge through that)

In the case of *jouhou*, we use *shirase* (news) and *chishiki* (knowledge) as Senses 1 and 2.

C. Experimental results

We performed word sense disambiguation in the four ambiguous words, *imi*, *kodomo*, *hoka*, and *jouhou*. Tables IV to VII show the number of data items in *imi*, *kodomo*, *hoka*, and *jouhou*. *imi* has three senses. *kodomo*, *hoka*, and *jouhou* have two senses each.

We show the experimental results of word sense disambiguation using three kinds of training datasets in Tables VIII and IX. In the tables, "Highest frequency" indicates the method that outputs a category with the highest frequency in an training dataset of SemEval 2, which is often used as a baseline method. "SemEval 2" indicates the method of using only a training dataset in SemEval 2. "SemEval 2 and paraphrasing" indicates the method of using both a training dataset in SemEval 2 and an additional constructed training dataset based on paraphrasing. "Paraphrasing" indicates the method of using only an additional constructed training dataset

TABLE VIII: Accuracies when using the maximum entropy method

Methods	Accuracies				
	imi (meaning)	*kodomo* (child)	*hoka* (other)	*jouhou* (information)	Total
Highest frequency	0.54 (27/50)	0.36 (18/50)	1.00 (50/50)	0.84 (42/50)	0.68 (137/200)
SemEval 2	0.50 (25/50)	0.56 (28/50)	1.00 (50/50)	0.86 (43/50)	0.73 (146/200)
SemEval 2 and paraphrasing	0.62 (31/50)	0.64 (32/50)	0.66 (33/50)	0.82 (41/50)	0.68 (137/200)
Paraphrasing	0.60 (30/50)	0.68 (34/50)	0.52 (26/50)	0.80 (40/50)	0.65 (130/200)
SemEval 2 and paraphrasing + changed	0.60 (30/50)	0.64 (32/50)	1.00 (50/50)	0.84 (42/50)	0.77 (154/200)
Paraphrasing + changed	0.54 (27/50)	0.66 (33/50)	1.00 (50/50)	0.84 (42/50)	0.76 (152/200)

TABLE IX: Accuracies when using the support vector machine

Methods	Accuracies				
	imi (meaning)	*kodomo* (child)	*hoka* (other)	*jouhou* (information)	Total
Highest frequency	0.54 (27/50)	0.36 (18/50)	1.00 (50/50)	0.84 (42/50)	0.68 (137/200)
SemEval 2	0.56 (28/50)	0.54 (27/50)	1.00 (50/50)	0.88 (44/50)	0.74 (149/200)
SemEval 2 and paraphrasing	0.58 (29/50)	0.68 (34/50)	0.66 (33/50)	0.82 (41/50)	0.68 (137/200)
Paraphrasing	0.68 (34/50)	0.56 (28/50)	0.66 (33/50)	0.82 (41/50)	0.67 (135/200)
SemEval 2 and paraphrasing + changed	0.60 (30/50)	0.64 (32/50)	1.00 (50/50)	0.86 (43/50)	0.77 (155/200)
Paraphrasing + changed	0.60 (30/50)	0.72 (36/50)	1.00 (50/50)	0.82 (41/50)	0.78 (157/200)

based on paraphrasing. Furthermore, we show two additional experimental results of changing the number of data items for cases using an additional constructed training dataset based on paraphrasing. "+ changed" indicates the number of data items are changed.

D. Discussion on experimental results

Firstly, we discuss the accuracy rates for all four words.

From Tables VIII and IX, we found that the method of changing the number of data items is very important. Although the accuracies of "SemEval 2 and paraphrasing" and "Paraphrasing" when not changing the number of the data items at an additional constructed training dataset are lower than those of "SemEval 2," the accuracies of "SemEval 2 and paraphrasing + changed," and "Paraphrasing + changed" when changing the number of data items in an additional constructed training dataset are higher than those of the method of "SemEval 2." We found that changing the number of data items and using an additional constructed training dataset were effective.

When we used the methods of changing the number of data items and using an additional constructed training dataset, the accuracy rates were nearly the same for the maximum entropy method and the support vector machine. When we used the maximum entropy method and the support vector machine, the accuracy rates were nearly the same as for the method of "SemEval 2 and paraphrasing" and the method of "Paraphrasing."

In the experiments, the support vector machine for using "Paraphrasing + changed" obtained the highest accuracy (0.78). We found that even if we used the method of using only an additional constructed training dataset based on paraphrasing where the number of data items is changed (we did not use a training dataset in SemEval 2), we could solve word sense disambiguation to some extent.

We examined the accuracy rates of each word. In the case of *imi* and *kodomo*, the accuracy rate of "SemEval 2" was low.

TABLE X: Features and α values for *kodomo* (child) when using an additional constructed training dataset

Sense 1 (*jidou* (pupil))		Sense 2 (*musuko* (son))	
Feature	normalized α values	Features	normalized α values
F1: *shisetsu* (establishment or school)	0.70	F1: *fuufu* (married couple)	0.65
F1: *shakai* (community)	0.55	F1: *kekkon* (marriage)	0.53

As in such a case, when the accuracy rate of "SemEval 2" was low, the accuracy rate of "Paraphrasing + changed" was rather higher. In the case of *jouhou*, the accuracy rate of "SemEval 2" was high. As in such a case, when the accuracy rate of "SemEval 2" was high, the accuracy rate of "Paraphrasing + changed" was comparatively rather low. In the case of *hoka*, the accuracy rate of "SemEval 2" was nearly the same as the accuracy rate of "Paraphrasing + changed."

E. Discussion based on feature analysis

In the maximum entropy method, a more important feature has a higher α value. A normalized α value represents the degree of importance of the corresponding feature used to estimate a sense using the maximum entropy method. Normalized α values range from zero to one. Further details can be found in the literature [12]. Features and α values for *kodomo* (child) when using increased training data items are shown in Table X.

shisetsu (establishment or school) and *shakai* (community) are useful for estimating Sense 1 (*jidou* (pupil)). *fuufu* (married couple) and *kekkon* (marriage) are useful for estimating Sense 2 (*musuko* (son)). When we used an additional constructed training dataset based on paraphrases, we obtained such features with high α values and obtained better accuracies by using the features than when we did not use an additional constructed training dataset. This indicates that an additional training dataset based on paraphrases is effective.

We checked features and α values in the case of not using an additional constructed training dataset. In this case, *shisetsu* (establishment or school), *shakai* (community), *fuufu* (married couple) and *kekkon* (marriage) had low α values, such features were not effectively used in estimation, and the accuracies became lower. This was caused by the fact that when not using an additional constructed training dataset, the number of the data items in a training dataset was small.

IV. Related Studies

Mihalcea et al. [2] proposed a method for generating a sense-tagged corpus gathering sentences including words that are automatically extracted from definitions in a word dictionary. Agirre et al. [13] conducted experiments using Mihalcea et al.'s method and the decision list method and reported that the results were worse than the results obtained using hand-tagged corpora. They suggested that the automatically-tagged corpora would provide misleading features. In contrast to their studies, we used a method for generating a sense-tagged corpus gathering sentences including words that we selected manually. Our study would be unlikely to provide misleading features, because we selected words used for gathering sentences manually. Furthermore, the maximum entropy method and the support vector machine we used in our experiments are generally likely to obtain higher accuracies than the decision list method.[2]

V. Conclusions

In this study, we performed word sense disambiguation using machine learning. We increased the data items of a training dataset by semi-automatically constructing data items of a training dataset using paraphrases of an ambiguous word. We performed word sense disambiguation using machine learning with the additional constructed training dataset.

In the accuracy rates of all four words, the support vector machine of using "only an increased training dataset based on paraphrasing where the number of data items is changed" obtained the highest accuracy (0.78). We found that even if we used "only an additional constructed training dataset based on paraphrasing where the number of data items is changed" (we did not use a training dataset in SemEval 2), we could solve word sense disambiguation to some extent.

We found that the method of changing the number of data items is very important. Although the accuracies of using an additional constructed training dataset based on paraphrasing are lower than those of using only an original training dataset, the accuracies of an additional constructed training dataset based on paraphrasing where the number of data items is changed are higher than those of using only the original training dataset.

In the case of the words, *imi* and *kodomo*, where using "only a training dataset in SemEval 2" obtained a low accuracy, using "only an additional constructed training dataset based on paraphrasing where the number of data items is changed" was better than using "only a training dataset in SemEval 2." In the case of the word, *jouhou*, where using "only a training dataset in SemEval 2" obtained a high accuracy and using "only an additional constructed training dataset based on paraphrasing where the number of data items is changed" was worse than using "only a training dataset in SemEval 2."

By using normalized α values in the maximum entropy method, we checked useful features. We found that when we used an additional constructed training dataset, some features were effective for sense estimation, and when we did not use an additional constructed training dataset, some features were not effective for sense estimation. This indicates that an additional constructed training dataset is effective.

In the future, we would like to perform experiments by increasing the number of words used in the experiments.

[2]In our experiments, we confirmed that the decision list method obtained lower accuracies than the maximum entropy method and the support vector machine.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 26330252.

References

[1] H. Shinnou, M. Murata, K. Shirai, F. Fukumoto, S. Fujita, M. Sasaki, K. Komiya, and T. Inui, "Classification of word sense disambiguation errors using a clustering method," *Journal of Natural Language Processing*, vol. 22, no. 5, pp. 319–362, 2015, (in Japanese).

[2] R. Mihalcea and D. I. Moldovan, "An automatic method for generating sense tagged corpora," in *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, 1999, pp. 461–466.

[3] M. Murata, M. Utiyama, K. Uchimoto, Q. Ma, and H. Isahara, "Japanese word sense disambiguation using the simple bayes and support vector machine methods," *Proceedings of SENSEVAL-2*, pp. 135–138, 2001.

[4] E. S. Ristad, "Maximum Entropy Modeling for Natural Language," ACL/EACL Tutorial Program, Madrid, 1997.

[5] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, 2000.

[6] T. Kudoh, "TinySVM: Support Vector Machines," http://cl.aist-nara.ac.jp/ taku-ku// software/TinySVM/ index.html, 2000.

[7] M. Okumura, K. Shirai, K. Komiya, and H. Yokono, "On semeval-2010 japanese wsd task," *Journal of Natural Language Processing*, vol. 18, no. 3, pp. 293–307, 2011.

[8] M. Murata, M. Kojima, T. Minamiguchi, and Y. Watanabe, "Automatic selection and analysis of Japanese notational variants on the basis of machine learning," *International Journal of Innovative Computing, Information and Control*, vol. 9, no. 10, pp. 4231–4246, 2013.

[9] M. Murata, Y. Goda, and M. Tokuhisan, "Automatic selection and analysis of synonyms in japanese sentences using machine learning," in *The 2014 International Conference on Artificial Intelligence*, 2014, pp. 1–4.

[10] M. Murata, K. Kanzaki, K. Uchimoto, Q. Ma, and H. Isahara, "Meaning sort — three examples: dictionary construction, tagged corpus construction, and information presentation system —," *Computational Linguistics and Intelligent Text Processing, Second International Conference, CICLing 2001, Mexico City, February 2001 Proceedings*, pp. 305–318, 2001.

[11] NLRI, *Bunrui Goi Hyou*. Shuuei Publishing, 1964.

[12] M. Murata, K. Uchimoto, M. Utiyama, Q. Ma, R. Nishimura, Y. Watanabe, K. Doi, and K. Torisawa, "Using the maximum entropy method for natural language processing: Category estimation, feature extraction, and error correction," *Cognitive Computation*, vol. 2, no. 4, pp. 272–279, 2010.

[13] E. Agirre and D. Martinez, "Exploring automatic word sense disambiguation with decision lists and the web," in *The COLING-2000 Workshop on Semantic Annotation and Intelligent Content*, 2000, pp. 11–19.