IPSJ SIG Technical Report
Vol.2013-NL-212 No.3
2013/7/18

Use of Combined Topic Models in Unsupervised Domain Adaptation for Word Sense Disambiguation

Shinya Kunii[1,a)]   Hiroyuki Shinnou[1,b)]

Abstract: Topic models can be used in the unsupervised domain adaptation for Word Sense Disambiguation (WSD). In the domain adaptation task, three types of topic models are available: (1) a topic model constructed from the source domain corpus: (2) topic model constructed from the target domain corpus, and (3) a topic model constructed from both domains. Basically, three topic features made from each topic model are added to the normal feature used for WSD. By using the extended features, SVM learns and it solves WSD. However, the topic features constructed from source domain has weights describing the similarity between the source corpus and the entire corpus. In six transitions of domain adaptation using three domains, we conducted experiments by varying the combination of topic features, and show the effectiveness of the proposed method.

Keywords: word sense disambiguation, domain adaptation, topic model, thesaurus, unsupervised learning

1. Introduction

In this paper, we propose an unsupervised method of domain adaptation for Word Sense Disambiguation (WSD) using topic models.

An inductive learning method is used in many tasks of natural language processing. In inductive learning, training data is created from corpus A, and a classifier learns from the training data. The original task is solved by using the classifier. During this analysis,data for the task is in corpus B, which differs from the domain of corpus A. In cases, the classifier learned from corpus A (i.e., the source domain) cannot analyze data corpus B (i.e., the target domain). This problem is called the domain adaptation problem, which is also regarded as a component of transfer learning in the field of machine learning. The domain adaptation problem has been extensively researched in recent years.

The method of domain adaptation can be divided into two groups from the viewpoint of whether labeled data is to be used in the target domain. When using labeled data, it is called supervised learning, while unsupervised learning does not use labeled data. There is substantial research on supervised learning techniques. Conversely,not much attention has been paid to unsupervised learning because of low precision; however, we adopt the unsupervised learning approach because it is does not require labeling.

Shinnou and Sasaki examined the unsupervised domain adaptation for WSD [17]. In his study, the topic model is built from the target domain corpus, and topic features constructed from this topic model are added to training data in both source and target domains. As a result, the accuracy of the classifier made by training data in the source domain is improved; however, in his study, the topic model is made by only the target domain. As indicated by Shinnou, it is unclear how topic models can be used for WSD. Further, in the domain adaptation task for WSD, the following three types of topic models are available: (1) a topic model constructed from the source domain corpus; (2) a topic model constructed from the target domain corpus, and (3) a topic model constructed from both domains. It is also unclear whether there is an effective combination of these topic models. The aim of this paper is to illuminate the latter problem.

The use of topic models in this paper adopts a similar approach to Shinnou [17]. Basically, three topic features made from each topic model are added to the normal features used for WSD, and the classifier learns using the extended features; however, the topic features constructed from the source domain has weights describing the similarity between the source corpus and the entire corpus because the topic features made from the source domain do not necessarily improve the accuracy of WSD, and sometimes actually reduce the accuracy. When it can be determined that a topic feature made from the source domain is effective for WSD, the value of r is approximately 1. In contrast, when it can be determined that a topic feature made from the source domain is not effective for WSD, the value of r is approximately 0.

The weight r is set by following equation:

$$r = \frac{KL(T, S+T)}{KL(T, S+T) + KL(S, S+T)}$$

where S is the source domain corpus, T is the target domain corpus, and S+T is the combined domain corpus; further, KL(A,B) is the Kullback Leibler (KL) divergence of A on criterion B.

In our experiments, we chose three domains, PB (books), OC (Yahoo! Chie Bukuro), and PN (news) in the BCCWJ corpus, and selected 17 ambiguous words that had a comparatively high

[1] 4-12-1 Nakanarusawa, Hitachi, Ibaraki 316–8511, Japan
[a)] 13nm707s@hcs.ibaraki.ac.jp
[b)] shinnou@mx.ibaraki.ac.jp

frequency of appearance in each domain.

Domain adaptation has the following six transitions: (1) from PB to OC, (2), from OC to PB, (3), from PB to PN, (4), from PN to PB, (5), from OC to PN, and (6) from PN to OC. In every domain adaptation, we conducted experiments by varying the combination of topic features. Through our experiments, we show the effectiveness of our proposed method.

2. Use of the Topic Model for WSD

In recent years, supervised learning approach have a great success for WSD, but this approach has the data sparseness problem. Generally, a thesaurus is used for the data sparseness problem. There are two types of the thesaurus which is constructed by hand and constructed automatically from a corpus. The former has a high quality, but has the domain dependence problem.The latter is not so high quality, and has an advantage that can be constructed form each the domain. In this paper, the latter is used in oder to deal with the domain adaption problem.

Topic model is a stochastic model that introduced K-dimensional latent topics z_i into generation of documents d.

$$p(d) = \sum_{i=1}^{K} p(z_i)p(d|z_i)$$

$p(w|z_i)$ for each word can be obtained by using Latent Dirichlet Allocation (LDA) [1],which is one of the topic models. Soft clustering can be done by using LDA and regarding the topic z_i as a cluster.

Suitable $p(w|z_i)$ for the domain is obtained by using the domain corpus and LDA. There are several studies [11][3][2] that use information of $p(w|z_i)$ for WSD, and Hard tagging approach [4] is used in this paper . Hard tagging approach is a method that give the word w to the topic of the highest relevance $z_{\hat{i}}$.

$$\hat{i} = \arg\max_{i} p(w|z_i)$$

First, when the number of topic is fixed K, a K-dimensional vector$t$ is prepared. Second,the topic of the highest relevance for each word $w_j (j = 1 \sim n)$ in a input example is evaluate ,and the value of $\hat{i}$-dimension on the vector t is set 1. Then, this operation proceed from w_1 to w_n. The vector made by this process is called topic features. The topic features made are added to the normal feature used for WSD, and extended features is used in learning and discrimination.

The normal features in this paper are the word in front of and behind the target word, part-of-speech in front of and behind the target word, and three content words in front of and behind the target word.

3. Three Types of Topic Features

In domain adaptation, the following three types of topic models are available: (1) a topic model constructed from the source domain corpus; (2) a topic model constructed from the target domain corpus, and (3) a topic model constructed from the both domains corpus. Three types of topic features can be made from three topic models.

The topic features made from the source domain is denoted by tp(S). The topic features made from the target domain is denoted by tp(T). The topic features made from the both domain is denoted by tp(S+T). The normal features used for WSD is denoted by B.

The following cases using the topic features for WSD are considered:
(1) B + tp(T)
(2) B + tp(S+T)
(3) B + tp(T) + tp(S+T)
(4) B + tp(T) + tp(S)
(5) B + tp(T) + tp(S+T) + tp(S)
(6) B + tp(T) + tp(S+T) + r * tp(S)

(1) and (2) are simply uses of the topic features for reflecting the knowledge of the target domain. (3) , which has the weight of the knowledge of the target domain,is also a promising method. A problem occurs that how tp(S) is used.

Currently, the key to a solution is how the knowledge of the source domain is used in domain adaptation. When the knowledge of the source domain is used, it does not necessarily improve the accuracy of WSD, and sometimes actually reduce the accuracy. Because of this, there is no guarantee that (4) is better than (1), (2) and (3).

(5) use tp(S), but is a promising method. This idea is similar to Daumé [5]. In study of Daumé, vector x_s of training data in the source domain is mapped to augmented input space $(x_s, x_s, 0)$, and vector x_t of test data in the target domain is mapped to augmented input space $(0, x_t, x_t)$. Classification problems are solved by using the augmented vector. This is known as the very simply and the high effectiveness method. This method is thought that an effect shows up in domain adaptation because the weight is learned by overlapping the characteristics common to the source and the target domain. It can be considered that (5) is added the knowledge tp(S+T) common to the knowledge of the source domain tp(S) and the knowledge of the target domain tp(T).

The proposed method in this paper is (6), and is the amended (5). As mentioned above,the weight has in (6) because the knowledge of the source domain tp(S) can have a bad influence on accuracy of WSD.

4. The Weight in the Source Domain

In this paper, the topic features are used as follows:

B + tp(T) + tp(S+T) + r * tp(S)

A problem occurs a apposite setting of the weight r.

It is considered that the weight r is the degree of the general knowledge which the source domain has.

Generally,in domain adaptation ,the key to the solution is how the knowledge of the source domain is used. This problem is closely related to the similarity of the source domain and the target domain.

4.1 Similarity Between Domains

In domain adaptation, it is necessary that the source domain is somewhat similar to the target domain. When the source domain is not similar to the target domain completely, it is clear that the source domain data is not useful in the target domain. It is difficult to define formally the degree of the similarity, and it is recognized one of the most important issue in domain adaptation since the dawn of domain adaptation.

Kamishima did not dare to give a concept of this similarity a universal definition, and did presuppose how the knowledge of the source domain is used in the target domain, and did point out that it is important how this assumption is modeled mathematically [7]. From this point of view, the similarity between the source and the target domains is measured, and it is normal to use the degree of this similarity for learning.

Asch measured the similarity among each the domain in part-of-speech tagging task, and showed that how the accuracy is reduced in domain adaptation by using the similarity [18]. Harimoto examined factors of performance decrement by varying the target domain in parsing [6]. Plank measured the similarity among each the domain in parsing, and chose the most suitable source domain in oder to analyze the target domain [14]. Ponomareva [15] and Remus [16] used the similarity among the domains for parameter on learning in sentiment classification. Those studies measured the similarity for every task. It is thought that the similarity among the domains depend on the target words in WSD. Komiya changed the learning methods for each target word by using the property*1 including the distance between domains [9] [8] [10].

4.2 Setting of the weight r

Measuring between the source and the target domains is mean that separating the common knowledge of the both domains and the specific knowledge because the similarity is intrinsically measured by comparing the common and the specific knowledge.

The weight r is considered to be the degree of the general knowledge that the source domain has. Because of this, it is important to how the general knowledge is set for calculating the weight r. The general knowledge is expressed by the combined domain corpus, that is contracted the the source and the target domain corpus. By combining two corpus, weights of the common part in two corpus is increased, and it is thought that the combined domain corpus approximates to the common part. By using KL divergence, $KL(S, S+T)$ is the distance between Corpus S and the general knowledge, and $KL(T, S+T)$ is the distance between Corpus T and the general knowledge. The following relationship is assumed:

$$r : 1 - r = KL(S, S+T) : KL(T, S+T)$$

By this assumption, r is calculated by the following equation:

$$r = \frac{KL(T, S+T)}{KL(T, S+T) + KL(S, S+T)}$$

Here, how to measure $KL(S, S+T)$ is describe in the following. Frequency of the nouns w in the corpus $S+T$ and in the corpus S is checked. The definition of $KL(S, S+T)$ is the following equation:

$$KL(S, S+T) = \sum_{w} p_s(w) \log \frac{p_s(w)}{p_{s+t}(w)}$$

where $p_{s+t}(w)$ is an occurrence probability in the corpus $S+T$, and is the following equation:

$$p_{s+t}(w) = \frac{f_{s+t}(w)}{N_{s+t}}$$

where $N_{s+t} = \sum_w f_{s+t}(w)$. $p_s(w)$ is an occurrence probability of the words w in the corpus S, and is defined by the following equation:

$$p_s(w) = \frac{f_s(w) + 1}{N_s + V}$$

where $N_s = \sum_w f_s(w)$, and V is the number of types of nouns in the corpus $S+T$.

5. Experiments

In our experiments, we chose three domains, PB (books), OC (Yahoo! Chie Bukuro), and PN (news) in the BCCWJ corpus [12], and selected 17 ambiguous words that had a comparatively high frequency of appearance in each domain. Table1*2 shows words and the number of word sense on dictionary in our experiments. PB and OC corpus are gotten from BCCWJ corpus, and PN is gotten from Mainichi newspaper in 1995.

Table 1 Target words

word	PB freq. of word	PB # of senses	OC freq. of word	OC # of senses	PN freq. of word	PN # of senses
言う (iu)	1114	2	666	2	363	2
入れる (ireru)	56	3	73	2	32	2
書く (kaku)	62	2	99	2	27	2
聞く (kiku)	123	2	124	2	52	2
来る (kuru)	104	2	189	2	19	1
子供 (kodomo)	93	2	77	2	29	2
時間 (jikan)	74	2	53	2	59	2
自分 (jibun)	308	2	128	2	71	2
出る (deru)	152	3	131	3	89	3
取る (toru)	81	7	61	7	43	7
場合 (bai)	137	2	126	2	73	2
入る (hairu)	118	4	68	4	65	3
前 (mae)	160	2	105	3	106	4
見る (miru)	273	6	262	5	87	3
持つ (motu)	153	3	62	4	59	3
やる (yaru)	156	4	117	3	27	2
ゆく (yuku)	133	2	219	2	27	2
Average	193.9	2.94	150.6	2.88	72.2	2.59

We conduct six transitions since there are three domains. We conducted experiments by varying the combination of the topic features (as mentioned section 3) for above target words on each method, and obtained the average accuracy rate for the words.

Topic model learned by using LDA *3, and the number of topics was fixed 100. Table2 shows the result of our experiments.

The accuracy rate of method that does not use topic model is lower than the other, and showed the effectiveness of topic model for WSD. The proposed method (7) is the highest accuracy rate, and showed the effectiveness.

6. Discussions

6.1 Use of the Topic Model

In this paper, the topic features are made from topic models, and added to the normal features. Several uses of the topic model for WSD have been suggested.

Use of the topic model for WSD can be divided into the direct and the indirect use.

*1 All those property can be called the similarity among the domains

*2 word sense is underlain the Iwanami Kokugo Jiten in the Japanese dictionary and middle level sense is targeted in our experiments. 「入る (hairu)」 is defined three word sense in the dictionary, but is defined four word sense in PB and PB because a novel sense of the word appears in BCCWJ corpus.

*3 `http://chasen.org/~daiti-m/dist/lda/`

Table 2 Experimental result (average accuracy rate %)

	OC→PB	OC→PN	PB→OC	PB→PN	PN→OC	PN→PB	Average
(1) B	74.18	70.18	70.38	76.94	69.25	74.88	72.64
(2) B + tp(T)	74.58	68.40	70.89	77.78	70.13	75.80	72.93
(3) B + tp(S+T)	73.48	70.46	72.70	78.50	70.25	76.24	73.61
(4) B + tp(T) + tp(S+T)	73.61	69.88	72.45	78.90	70.36	76.86	73.68
(5) B + tp(T) + tp(S)	73.61	68.79	72.09	78.91	70.17	76.48	73.34
(6) B + tp(T) + tp(S+T) + tp(S)	73.92	68.70	72.18	79.41	70.53	76.71	73.58
(7) B + tp(T) + tp(S+T) + r*tp(S) (proposed method)	73.63	69.89	72.14	79.08	70.58	77.17	73.75
Weight r	0.0174	0.01139	0.9825	0.35655	0.98861	0.6434	

The indirect use is to fortify the resource used for WSD. Cai used Bayesian Network for WSD, and improved the original Bayesian Network by innovating the topic features made from topic model to Bayesian Network [4]. Boyd-Graber introduced the word sense of WordNet as the additional latent variable into LDA, and used topic model to search synset from WordNet [3]. Li proposed a method of constructing a probability model for WSD depending on three circumstances, which Prior probability distribution of word sense was obtained from the corpus or not and the resource of paraphrase in corpus lacked [11].

The direct use is directly using the topic features made from topic model for WSD. The proposed method belongs to this type. Boyd-Graber estimated marginal probability distribution of the word using LDA, and estimated word sense form the probability distribution [2]. However, due to unsupervised learning,the normal features is not used for WSD, and it is not study that improve a classifier made from supervised learning by using topic model. Cai's paper described above, a method that the topic features are added to the normal features was implemented as a comparison method with the proposed method [4]. Cai conducted two experiments, which hard tag was a method that give the word w to the topic of the highest relevance, and soft tag was a method that use all topic of relevance, and pointed out that the soft tag is better.

From the viewpoint of easiness of implement, the direct use is better; however, in this case, the corpus domain which builds topic model, the size of the corpus and the number of topic have a great influence for the accuracy, and it is necessary to estimate the value of those. Especially, the corpus used in our experiments was 26.8MB in PB, was 0.4MB in OC and was 52.4MB in PN. The size of OC was smaller than the other. Therefore, the similarity between the OC and other was so small. When the source domain was OC, the weight r was also small.

6.2 Comparison with Existing Thesaurus

In this paper, topic models were used as thesaurus. We compared the proposed method and the the method that use existing thesaurus. We used Bunrui-goi-hyou[*4] as Existing thesaurus. Table3 shows the result.

The accuracy rate of the method that use topic models is higher than using existing thesaurus.

This result suggest that it is better to use topic models constructed form the corpus of domain that is targeted in the task than to use existing thesaurus when solving WSD. Moreover, considering this result, use of a combination of topic models and existing thesaurus can have a effectiveness. This point is for further study.

Table 3 Comparison with existing thesaurus

	the propose method	B + thesaurus
OC→PB	73.63	72.85
OC→PN	69.89	70.64
PB→OC	72.14	70.68
PB→PN	79.08	78.13
PN→OC	70.58	69.72
PN→PB	77.17	75.87
Average	73.75	72.98

*4 Japanese standard thesaurus

6.3 Domain Dependence of Thesaurus

When considering a domain adaptation problem, there is an idea that can use the common knowledge constructed form all domains for all domains in common. In fact, there are such tasks. For example, Mori improved the accuracy using the labeled data of each domain, and pointed outs that it is better to use the labeled data of all domains than using the labeled data of each domain[13].

In the task in this paper, if the combined corpus of all domains is made and the topic model is made from this the corpus, it is thought that the topic model can be used in each domain. This idea is the method (3) ,B + tp(S+T), achieved good evaluation value in the experiments results. Moreover, it is clear that the knowledge of the target domain has a effectiveness in the target domain, and it can be envisioned that the method (4),B + tp(T) + tp(S+T) ,has a effectiveness rather than the method (3), and the experiments results shows also that.

The problem is the way of using tp(S). Basically, tp(S) need not to be used; however, when the source domain corpus S is similar to the combined corpus S+T, the topic feature tp(S) has benefit in domain adaptation. In particular, when $KL(S, S+T)$ is only bigger than $KL(T, S+T)$, the topic feature tp(S) has benefit in domain adaptation. In this paper, tp(S) that has the weight is used, but actually tp(S) is also used as $r = 1$ only if the above and is not used in the other cases.

6.4 Domain Dependence of Thesaurus of Each Target Word

The weight r of tp(S) on the proposed method in this paper was set for each domain. There is an idea that the optimum method of domain adaptation for each word is different. We examined that whether use of the topic models differs for each word.

Table4 shows the method of the highest accuracy rate in domain adaptation for each word. In addition, the number of table4 corresponds to the number of methods, table2

Seen Table4, several words have the effective methods regard-

Table 4 the best method of each word

word	OC→PB	OC→PN	PB→OC	PB→PN	PN→OC	PN→PB
言う (iu)	1	2	3	1	6 7	3 5
入れる (ireru)	2	5	4	6	3	7
書く (kaku)	5	3	1 2 3 4 5 6 7	1 2 3 4 5 7	2	3 5 6 7
聞く (kiku)	6	4 7	3	2	2 4	3
来る (kuru)	3 4	1 2 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
子供 (kodomo)	5	1 2 3 5 6	4	4 7	4	3
時間 (jikan)	2 6	6	1 2 3 4 5 6 7	6	2 4 5 6 7	3
自分 (jibun)	4	1	4	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
出る (deru)	2	3 4 7	6	2 3 4	5	4
取る (toru)	1 2 4 5 6	4 7	3	6	5	2
場合 (bai)	1 3 4 6	1	2	1	3 6 7	3
入る (hairu)	4	1	5	6	3 5 6 7	7
前 (mae)	4	1 3	1	5 6	6 7	6
見る (miru)	1	1	1 3	1	3	2
持つ (motu)	1	2 6	3	3	2 3 4	1 2 3 4 5 6 7
やる (yaru)	1 2 3 5	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
ゆく (yuku)	4	1 2 3 4 5 6 7	4 6 7	1 3 4 5 6 7	1 2 3 4 5 6 7	2 4

less of the combination of the domains. For example, method (4) is better in word 「ゆく (yuku)」 and 「自分 (jibun)」,and method (5) is better in word 「書く (kaku)」. 「やる (yaru)」 and 「来る (kuru)」 do not depend substantially on the methods, and the other words do not depend on the certain method. Table4 also shows that the effective methods depends on the domains. In other words, it is thought that the effective use of the topic models in domain adaptation for WSD is determined from the target words and the domains.

7. Conclusions

In this paper, we proposed an unsupervised method of domain adaptation for word sense disambiguation using topic models. Concretely, each topic model is constructed form the source domain corpus, the target domain corpus and the both domain corpus.The topic features are made by each topic model. Therefore. three topic features are available. Three topic features made from each topic model are added to the normal features, and the extended feature are used in learning for WSD. However, regarding the topic features made from the source domain, this topic features have the weight because this topic features reduces the accuracy of BSD. This weight is obtained from the similarity between the two domains, and the similarity is measured by Kullback-Leibler divergence. In our experiments, we chose three domains, and selected 17 ambiguous words that had a comparatively high frequency of appearance in each domain. In every domain adaptation, we conducted experiments by varying the combination of topic features, and estimated the average accuracy rate of WSD. Eventually,the effectiveness of the proposed method is showed. In future, we will examine the more effective use of the topic models in the WSD task.

References

[1] Blei, D. M., Ng, A. Y. and Jordan., M. I.: Latent dirichlet allocation, *Machine Learning Reseach*, Vol. 3, pp. 993–1022 (2003).

[2] Boyd-Graber, J. and Blei, D.: Putop: Turning Predominant Senses into a Topic Model for Word Sense Disambiguation, *SemEval-2007*.

[3] Boyd-Graber, J., Blei, D. and Zhu, X.: A Topic Model for Word Sense Disambiguation, *EMNLP-CoNLL-2007*, pp. 1024–1033 (2007).

[4] Cai, J. F., Lee, W. S. and Teh, Y. W.: Improving Word Sense Disambiguation using Topic Features, *EMNLP-CoNLL-2007*, pp. 1015–1023 (2007).

[5] Daumé III, Hal: Frustratingly Easy Domain Adaptation, *ACL-2007*, pp. 256–263 (2007).

[6] Harimoto, K., Miyao, Y., and Tsujii, J.: Kobunkaiseki no bunyatekiou ni okeru seido teika youin no bunseki oyobi bunyakan kyori no sokutei syuhou (in Japanese), *The 16th Annual Meeting on Journal of Natural Language Processing*, pp. 27–30 (2010).

[7] Kamishima, T.: Transfer Learning (In Japanese), *The Japanese Society for Artificia Intelligence*, Vol. 25, No. 4, pp. 572–580 (2010).

[8] Komiya, K. and Okumura, M.: Automatic Determination of a Domain Adaptation Method for Word Sense Disambiguation using Decision Tree Learning, *IJCNLP-2011*, pp. 1107–1115 (2011).

[9] Komiya, K. and Okumura, M.: Automatic Domain Adaptation for Word Sense Disambiguation Based on Comparison of Multiple Classifiers, *PACLIC-2012*, pp. 75–85 (2012).

[10] Komiya, K. and Okumura, M.: Automatic selection of domain adaptation method for WSD using decision tree learning (In Japanese), *Journal of NLP*, Vol. 19, No. 3, pp. 143–166 (2012).

[11] Li, L., Roth, B. and Sporleder, C.: Topic Models for Word Sense Disambiguation and Token-based Idiom Detection, *ACL-2010*, pp. 1138–1147.

[12] Maekawa, K.: Design of a Balanced Corpus of Contemporary Written Japanese, *Symposium on Large-Scale Knowledge Resources (LKR2007)*, pp. 55–58 (2007).

[13] Mori, S.: Domain Adaptation in Natural Language Processing (in Japanese), *The Japanese Society for Artificia Intelligence*, Vol. 27, No. 4, pp. 365–372.

[14] Plank, B. and van Noord, G.: Effective measures of domain similarity for parsing, *ACL-2011*, pp. 1566–1576 (2011).

[15] Ponomareva, N. and Thelwall, M.: Which resource is best for cross-domain sentiment analysis?, *CICLing-2012* (2012).

[16] Remus, R.: Domain Adaptation Using Domain Similarity- and Domain Complexity-based Instance Selection for Cross-domain Sentiment Analysis, *Proceedings of the 2012 IEEE 12th International Conference on Data Mining Workshops (ICDMW 2012) Workshop on Sentiment Elicitation from Natural Text for Information Retrieval and Extraction (SENTIRE)*, pp. 717–723 (2012).

[17] Shinnou, H. and Sasaki, M.: Domain Adaptation for Word Sense Disambiguation using k-Nearest Neighbors Method and Topic Model (In Japanese), pp. NL–211 (2013).

[18] Van Asch, V. and Daelemans, W.: Using domain similarity for performance estimation, *Proceedings of the 2010 Workshop on Domain Adaptation for Natural Language Processing*, pp. 31–36 (2010).