Neural Machine Translation from Historical Japanese to Contemporary Japanese Using Diachronically Domain-Adapted Word Embeddings

Masashi Takaku Tosho Hirasawa Mamoru Komachi Kanako Komiya

Ibaraki University
4-12-1, Nakanarusawa, Hitachi, Ibaraki, 316-8511, Japan
{16t4027n, kanako.komiya.nlp}@vc.ibaraki.ac.jp

Tokyo Metropolitan University
6-6 Asahigaoka, Hino, Tokyo, 191-0065, Japan
{hirasawa-tosho@ed., komachi@}tmu.ac.jp

Abstract

This paper describes the first trial of neural machine translation (NMT) from historical Japanese to contemporary Japanese. To compensate for the lack of parallel data, we used pre-trained word embeddings for the input of the system and performed diachronic domain adaptation in the order of time. We investigated and compared an NMT system without pre-trained word embeddings, an NMT system with pre-trained word embeddings trained with contemporary Japanese, an NMT system with word embeddings diachronically domain-adapted at one time, and NMT systems with word embeddings that were gradually domain-adapted in the order of time. Although our system did not outperform statistical machine translation, experiments revealed that diachronic domain adaptation is effective, especially if it is performed in the order of time.

1 Introduction

This paper describes a neural machine translation (NMT) system for translation from historical Japanese to contemporary Japanese. In recent years, machine translation using deep learning, or NMT, has been intensively studied. Although there is a study of statistical machine translation (SMT) from historical Japanese to contemporary Japanese (Hoshino et al., 2014), to the best of our knowledge, there have been no studies on an NMT system that translates historical Japanese to contemporary Japanese. NMT systems generally output fluent translations. Therefore, NMT is expected to improve the fluency of contemporary Japanese translation. However, it is difficult to obtain a model with high performance when only small parallel corpora are available, as NMT systems usually require large parallel corpora for training (Koehn and Knowles, 2017). Because the available parallel corpus of historical and contemporary Japanese is small, NMT is not appropriate for translation from historical Japanese to contemporary Japanese.

To improve translation performance, translation models are sometimes initialized with pre-trained word embeddings trained with a large corpus for language pairs that do not have sufficient parallel corpora (Qi et al., 2018). We believe that this method can also be effective for the translation model from historical Japanese to contemporary Japanese.

To obtain high-quality word embeddings, it is desirable to train them using a large training corpus. Therefore, the use of word embeddings trained with a contemporary Japanese corpus that is much larger than the historical Japanese corpus is expected. However, when the word embeddings trained with contemporary Japanese are directly used for the translation model, the model is expected to have poorer performance because the domains of the word embeddings and inputs differ from each other.

In addition, because the parallel corpus of historical and contemporary Japanese contains literature from different time periods, the meaning of the words or the words themselves may change according to the period. We thus propose a method for initializing the translation model with word embeddings generated from a contemporary Japanese corpus and perform gradually diachronic domain adaptation by fine-tuning using the corpus written in each period in the order of time (see Section 3). We compared other initialization methods in which pre-trained word embeddings were directly used or were fine-tuned using the entire historical corpus at one time (see Section 4).

The findings of this paper are listed as follows:

(1) The bilingual evaluation understudy (BLEU) score of the proposed method outperformed that of other methods, as discussed in Section 5,

(2) The quality of translation displayed improvement when diachronically domain-adapted word embeddings in the order of time were used; the model could translate words that appeared only in a corpus of a specific time period (see Section 6),

(3) However, the translation model that was diachronically domain-adapted up to a certain period did not always exhibit the best translation performance on the test set of that period, as discussed in Section 6.

2 Related Work

Hoshino et al. (2014) proposed a method for obtaining a sentence-based parallel corpus using a rule-based score function for aligning sentences from a paragraph-based parallel corpus of historical and contemporary Japanese. They translated historical Japanese to contemporary Japanese using a SMT system trained with the corpus obtained by the proposed method, and demonstrated the effectiveness of their proposed sentence aligning method. We believe that our study was the first to utilize NMT for translation from historical Japanese to contemporary Japanese.

Much research has been conducted on natural language processing using word embeddings, or distributed representations. Classification tasks, or sequence labeling tasks, using deep learning usually utilize pre-trained word embeddings; however, pre-trained word embeddings are rarely used in NMT. This is because the translation model itself learns suitable word embeddings when it is trained with a large parallel corpus. However, the initialization of inputs with pre-defined word embeddings offers the potential for improved translation performance when the translation model is trained with only a small parallel corpus. Qi et al. (2018) demonstrated that the use of pre-trained word embeddings for the training of NMT improved the translation performance for language pairs that had only a small parallel corpus.

For the domain adaptation method of word embeddings, we employed fine-tuning. Faruqui et al. (2015) proposed the retrofitting method, demonstrating that fine-tuning with another corpus improved the quality of the word embeddings. Yaginuma et al. (2018) performed word sense disambiguation in Japanese using fine-tuned word embeddings.

In addition, Kim et al. (2014) performed diachronic fine-tuning. They automatically detected changes in language over time through a chronologically trained neural language model. They obtained word embeddings specific to each year and demonstrated that some words had changed their meanings. Based on their research, we believe that diachronically domain-adapted word embeddings can capture changes in language meanings over time.

3 Diachronic Domain Adaptation of Word Embeddings Using Historical Corpus

In this study, we propose the initialization of inputs to the NMT model with diachronically domain-adapted word embeddings. Following the study by Kim et al. (2014), we chronologically fine-tuned the word embeddings starting from the newest corpora (contemporary Japanese corpus) to the oldest corpora. This is the domatin adaptation of time.

We fine-tuned the word embeddings in the order of time to minimize the drift in meaning over time. We employed the fine-tuning method used by Yaginuma et al. (2018).

We used the parallel corpus of historical and contemporary Japanese from four periods: the modern period (after the Edo period), Muromachi period, Kamakura period, and Heian period [1].

The procedures of diachronic domain adaptation are as follows:

[1]Edo, Muromachi, Kamakura, Heian, and Nara Periods are from 1603 to 1868, from 1336 to 1573, from 1185 to 1333, from 794 to 1185, and from 710 to 794, respectively. These periods are defined according to the political systems by historians.

(1) Fine-tune word embeddings pre-trained with contemporary Japanese using the modern corpus,

(2) Fine-tune the corpus obtained in step (1) using the Muromachi corpus,

(3) Fine-tune the corpus obtained in step (2) using the Kamakura corpus,

(4) Fine-tune the corpus obtained in step (3) using the Heian corpus.

4 Experiments

4.1 Model

In our experiments, we employed an encoder-decoder model[2] based on long short-term memory (LSTM) with attention. OpenNMT[3], an open-source NMT tool, was used for implementation. We utilized two unidirectional LSTM layers for the hidden layers and global attention (Luong et al., 2015) for attention. The vector sizes of the word embeddings and the hidden layers were set to 200 and 512, respectively, for both the encoder and decoder. Adam was used as the optimization algorithm, and the learning rate was set to 0.001. The vocabulary size treated by the model was limited to 20,000 for each of the source and target data, and unknown words were processed as <unk> tokens. The hyper parameters were determined according to preliminary experiments.

We initialized the weights of the word embedding layer of the translation model. In this study, the word embeddings pre-trained with the contemporary Japanese corpus were diachronically domain-adapted using the historical Japanese corpus, and used to initialize the weights of the word embedding layer of the encoder of the translation model. These word embeddings were also directly used for the decoder of the translation model.

We used the BLEU score (Papineni et al., 2002) to evaluate the translation model. Each method was given different seeds validated on each 5,000th step and was tested with the translation model with the highest BLEU score. The average BLEU scores over three trials using different seeds were evaluated as the scores of the translation models.

For comparison, we conducted experiments in which the weights of the word embedding layer were directly initialized with word embeddings pre-trained with the contemporary Japanese corpus without fine-tuning. In addition, we performed initialization with word embeddings pre-trained with the contemporary Japanese corpus and fine-tuned with the entire historical corpus at one time. Furthermore, we evaluated ensemble methods of models of diachronic domain adaptation at one time and models of diachronic domain adaptation in order of time. For the ensemble methods, we used the ensemble option of OpenNMT [4].

4.2 Data Set

We utilized a parallel corpus of historical and contemporary Japanese extracted by Hoshino et al. (2014) for translation. The sentences in this corpus were extracted from the corpus we used for fine-tuning: The Complete Collection of Japanese Classical Literature published by Shogakukan [5]. This corpus can be classified into five sub-corpora based on the periods in which each piece of literature was written. The statistics of the corpus are presented in Table 3. The periods in which each piece of literatures was written was determined by referring to the guide to the contents on the official website of the corpus [6].

Following Hoshino et al. (2014), we used three sub-corpora for the training and test data of the translation model. The modern Japanese corpus consisted of texts written after the Edo period, while the Kamakura corpus consisted of texts written in the Kamakura period. The Heian corpus consisted of texts written in the Heian period. The number of sentences in the modern, Kamakura, and Heian corpora were 4,577, 30,075, and 52,032, respectively, for a total of 86,684 sentences. The Muromachi corpus consisting of literature written in the Muromachi period was used only for fine-tuning because previ-

[2]We also tried a transformer model but the performance greatly varied depending on each trial. Also, the averaged performance did not surpass the encoder-decoder model. Therefore we decided to use an encode-decoder model.

[3]https://github.com/OpenNMT/OpenNMT

[4]https://github.com/OpenNMT/OpenNMT-py/pull/732

[5]https://japanknowledge.com/en/contents/koten/

[6]https://japanknowledge.com/en/contents/koten/title.html

ous research did not use it for testing. We did not use literature written in the Nara period at all because previous research did not use it for testing. Literature written in this period is also not suitable for fine-tuning because it is the oldest literature.

We divided the parallel corpus into training, development, and test corpora following previous research. The number of sentences were 82,591, 2,093, and 2,093, respectively (see Table 1). We randomly selected the test sentences. The ratio of the number of sentences of the entire corpus for each period in which the sentences were written is identical to that of the text examples. The number of sentences in the modern, Kamakura, and Heian corpora was 4577, 30,075, and 52,032, respectively, totaling to 86,684 sentences. Therefore, the number of sentences in the modern, Kamakura, and Heian corpora was 123, 739, and 1231, respectively. (123:739:1,231) = (4,577:30,075:52,032)

The number of examples in the test set is presented in Table 2. We used MeCab v0.996[7] as a morphological analyzer and UniDic for Early Middle Japanese v1.3[8] (Ogiso et al., 2012) and UniDic v2.3.0[4] (Maekawa et al., 2010) as dictionaries for historical and contemporary Japanese, respectively. We limited the length of an input or output sentence to 100 words.

Historical Japanese	
Total Number of Sentences	86,684
Vocabulary Size	49,200
Number of Tokens	2,774,745
Contemporary Japanese	
Total Number of Sentences	86,684
Vocabulary Size	45,690
Number of Tokens	3,611,783

Table 1: Parallel corpus of historical and contemporary Japanese.The data for translation are extracted from parallel corpus of Complete Collection of Japanese Classical Literature (see Table 3).

4.3 Word Embeddings

We used NWJC2vec (Shinnou et al., 2017) for the word embeddings for contemporary Japanese. These word embeddings were generated from the NWJC-2014-4Q dataset (Asahara et al., 2014), which is an enormous Japanese corpus developed using the word2vec tool (Mikolov et al., 2013a; Mikolov et al., 2013b; Mikolov et al., 2013c). Tables 6 and 4 present summary statistics for the NWJC-2014-4Q data and the parameters used to generate the word embeddings, respectively.

[7]https://taku910.github.io/mecab/

[8]https://unidic.ninjal.ac.jp/

Period	Number of Example
Modern Test Set	123
Kamakura Test Set	739
Heian Test Set	1,231

Table 2: Number of test example according to period

CBOW or skip-gram	-cbow	1
Dimensionality	-size	200
Number of surrounding words	-window	8
Number of negative samples	-negative	25
Hierarchical softmax	-hs	0
Minimum sample threshold	-sample	1e-4
Number of iterations	-iter	15

Table 4: Parameters used to generate NWJC2vec

We followed Yaginuma et al. (2018) for the parameters for fine-tuning NWJC2vec (see Table 5). The other parameters were set to the default settings.

CBOW or skip-gram	-cbow	1
Dimensionality	-unit	200
Number of surrounding words	-window	5
Number of negative samples	-negative	5
Batch size	-batchsize	1000
Number of iterations	-iter	10

Table 5: Parameters used to fine-tune NWJC2vec

5 Results

Table 7 presents the BLEU scores of the entire test data according to each model. *SMT* (Hoshino et al., 2014) refers to the results of Hoshino et al. (2014), who used an SMT system to perform translation from historical Japanese to contemporary Japanese. *Baseline* refers to the results of the NMT model when only the parallel corpus of historical and con-

	Total Number of Sentences	Vocabulary Size	Number of Tokens
Modern	22,485	25,584	544,293
Muromachi	12,640	14,931	386,101
Kamakura	35,020	29,062	933,190
Heian	59,744	29,520	1,543,102
Nara	4,832	6,013	112,094
Total	134,721	61,345	3,518,780

Table 3: Parallel corpus of Complete Collection of Japanese Classical Literature published by Shogakukan

Number of URLs collected	83,992,556
Number of sentences (Some are overlapped)	3,885,889,575
Number of sentence (No overlapping)	1,463,142,939
Number of words (tokens)	25,836,947,421

Table 6: Statistics for the NWJC-2014-4Q dataset

temporary Japanese was used without using pre-trained word embeddings. *NWJC2vec* refers to the results of the NMT model when NWJC2vec was directly input to the system without fine-tuning. *Entire historical corpus* represents the results of the translation model when NWJC2vec was fine-tuned with the entire historical corpus at one time. *+Ensemble* refers to the results of an ensemble method. We evaluated two types of ensemble methods. The first, *+Ensemble at one time* is an ensemble method using the top three translation models of the *Entire historical corpus*. *(1) Modern*, *(3) Modern → Muromachi → Kamakura*, and *(4) Modern → Muromachi → Kamakura → Heian* are the results of translation models in which NWJC2vec was diachronically domain-adapted in the order of time. We did not evaluate (2) Modern → Muromachi because previous research did not use texts written in the Muromachi period for the training and test data. The second ensemble method, *+Ensemble in order of time*, is the ensemble method of models (1), (3), and (4).

According to Table 7, the best method among all NMT systems was the ensemble methods using diachronic domain adaptation in the order of time. Particularly, it outperformed the ensemble method of the top three models of diachronic domain adaptation using the entire contemporary Japanese corpus at one time. In addition, all diachronic domain adaptation methods in the order of time surpassed the *Entire historical corpus*. These results imply that diachronic domain adaptation in the order of time is effective.

Moreover, all methods using fine-tuning outperformed the baseline, although NWJC2vec, the model that directly used the word embeddings pre-trained with contemporary Japanese, were unable to outperform the baseline. This result indicates that fine-tuning using the historical corpus is effective. However, even the best NMT model did not surpass the SMT model. We believe that this is because NMT requires more parallel data than SMT. The BLEU score of the best NMT method was more than 2 points higher than that of the baseline method. However, the differences between models of diachronic domain adaptation in the order of time and models of diachronic domain adaptation at one time were not very large.

In addition, Table 7 indicates that the BLEU score decreased for earlier time periods when the word embeddings were diachronically domain-adapted in the order of time. Table 8 lists the BLEU scores of the proposed method according to each period. According to this table, the translation model in which (1) only the modern corpus was used for fine-tuning was the best for the modern test set. In addition, (4) diachronic domain adaptation up to the Heian period was the best for the Kamakura test set and outperformed (3) diachronic domain adaptation up to the Kamakura period. Furthermore, (1) diachronic domain adaptation up to the modern period was the best for the Heian test set. The second best was (3) diachronic domain adaptation up to the Kamakura

Method	BLEU
SMT (Hoshino et al., 2014)	28.02
Baseline	19.22
NWJC2vec	19.16
Entire historical corpus	19.24
+Ensemble at one time	20.94
Diachronic domain adaptation in order of time	
(1) Modern	**19.43**
(3) Modern → Muromachi → Kamakura	19.33
(4) Modern → Muromachi → Kamakura → Heian	19.29
+Ensemble in order of time	21.59

Table 7: BLEU scores of entire test data according to each model

period, whereas (4) diachronic domain adaptation up to the Heian period was the worst.

6 Discussion

Although the BLEU score did not significantly improve for the proposed method, some examples demonstrated the effectiveness of the proposed method.

Table 9 presents translation examples in which (a) diachronic domain adaptation up to the Kamakura period was better and (b) the ensemble method of the proposed methods was better.

For example, (a) "やまとうた (Japanese poems)" could not be translated to "和歌 (Japanese poem)" until only the modern corpus was used for fine-tuning; however, it could be translated correctly after the Kamakura corpus was used. In this example, diachronic domain adaptation improved the translation.

Furthermore, example (b) in Table 9 demonstrates that the word "騒がしき (noisy)" was erroneously translated when the word embeddings were fine-tuned with the entire data set at one time, but were correctly translated when they were gradually domain-adapted in the order of time. This example also demonstrates the effectiveness of the proposed method.

Next, we consider the effects of domain adaptation in the order of time. We hypothesized that the translation model would exhibit the best performance when the model was diachronically domain-adapted up to the time when the test set was written; however, Table 7 indicates that this hypothesis was not always correct. When the proposed method was used, unknown words decreased for earlier time periods because the fine-tuning method we used following Yaginuma et al. (2018) added a new entry when the new corpus included a new word that exceeded the threshold value. However, (1) diachronic domain adaptation up to the modern period was the best not only for the modern test corpus but also for the Heian corpus. However, the difference between the models for the Heian corpus was rather subtle compared to that for the modern corpus or Kamakura corpus; it was only 0.13 for the Heian corpus but 1.65 for the modern corpus and 0.6 for the Kamakura corpus. In other words, for the Heian corpus, there was no large difference based on the translation model of time.

Future work will include the investigation of effects of other word embeddings, such as fastText and Glove. In addition, diachronic domain adaptation using contextual word representations, such as ElMo and BERT, would be interesting.

7 Conclusion

This paper is the first to present an NMT system that translates historical Japanese to contemporary Japanese. We proposed diachronic domain adaptation of word embeddings in the order of time. We gradually fine-tuned the word embeddings for the input of the system in the order of time using a corpus written in each period. The NMT results were unable to surpass the results of the SMT system due to the lack of sufficient parallel data. In addition, the hypothesis that the translation model should have

	Modern	Kamakura	Heian
(1) Modern	5.24	25.16	19.53
(3) Modern → Muromachi → Kamakura	4.09	25.65	19.43
(4) Modern → Muromachi → Kamakura → Heian	3.59	25.76	19.40

Table 8: BLEU scores of the proposed method according to each period

(a) An example where diachronic domain adaptation until Kamakura period is better	
Input Sentence:	やまとうた (Japanese poems) の 道 、 浅き に 似 て 深く 、
Reference translation:	和歌 (Japanese poems) の 道 は 、 浅い よう で い て じつ は 深く 、
English translation:	The soul of Japanese poems seems shallow, but it is in fact profound.
Baseline:	<unk> の 道 は 、 浅い の に 似 て 深く 、
Modern:	<unk> の 道 、 浅い の と 同様 に 、深く て 、
Modern → Muromachi → Kamakura:	和歌 (Japanese poems) の 道 は 、 浅い 時代 に 似 て 深く 、
(b) An example where ensemble method of proposed methods is better	
Input sentence:	大饗 に 劣ら ず 、 あまり 騒がしき (noisy) まで なん 集ひ たまひ ける 。
Reference translation:	大 饗 の とき に 劣ら ない ほど 、 あまり に 騒がしい (noisy) まで 大勢 お 集まり に なる の だっ た 。
English translation:	There came together as many people as people at a royal party and it was almost too noisy.
Baseline:	大 饗 に も 劣ら ず 、 あまり に も あわただしい (hasty) くらい に お 集まり に なっ た 。
Entire historical corpus:	大 饗 に 負け ず 、 あんまり 暑い (hot) まで 集まっ て お いで に なっ た 。
Diachronic domain adaptation:	大 饗 に 劣ら ず 、 あまり に 騒がしい (noisy) まで 集まっ て お いで に なっ た 。

Table 9: Translation examples

the best performance when the input of the system is diachronically domain-adapted up to the period in whihc the test corpus is written was not always correct. However, the translation performance when the word embeddings were domain-adapted in the order of time was better than that when the embeddings were domain-adapted at one time. In addition, some examples in which diachronic domain adaptation improved the translations were observed.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 17H00917 and a project of the Center for Corpus Development, National Institute for Japanese Language and Linguistics. We would like to thank Hikaru Yokono, who gave us helpful advice and the data set used in previous research.

References

Masayuki Asahara, Kikuo Maekawa, Mizuho Imada, Sachi Kato, and Hikari Konishi. 2014. Archiving and analysing techniques of the ultra-large-scale web-based corpus project of NINJAL, Japan. *Alexandria*, 25(1-2):129–148.

Manaal Faruqui, Jesse Dodge, Sujay Kumar Jauhar, Chris Dyer, Eduard Hovy, and Noah A. Smith. 2015. Retrofitting word vectors to semantic lexicons. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1606–1615, Denver, Colorado, May–June. Association for Computational Linguistics.

Sho Hoshino, Yusuke Miyao, Shunsuke Ohashi, Akiko Aizawa, and Hikaru Yokono. 2014. Machine translation from historical Japanese to contemporary Japanese using parallel corpus. In *Proceedings of the NLP2014, (In Japanese)*, pages 816–819.

Yoon Kim, Yi-I Chiu, Kentaro Hanaki, Darshan Hegde, and Slav Petrov. 2014. Temporal analysis of language through neural language models. In *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science*, pages 61–65, Baltimore, MD, USA, June. Association for Computational Linguistics.

Philipp Koehn and Rebecca Knowles. 2017. Six challenges for neural machine translation. In *In Proceedings of the 2017 Annual Meeting of the Association for Computational Linguistics*, pages 28–39.

Thang Luong, Hieu Pham, and Christopher D. Manning. 2015. Effective approaches to attention-based neural machine translation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1412–1421, Lisbon, Portugal, September. Association for Computational Linguistics.

Kikuo Maekawa, Makoto Yamazaki, Takehiko Maruyama, Masaya Yamaguchi, Hideki Ogura, Wakako Kashino, Toshinobu Ogiso, Hanae Koiso, and Yasuharu Den. 2010. Design, compilation, and preliminary analyses of Balanced Corpus of Contemporary Written Japanese. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 2010)*, pages 1483–1486.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. Efficient estimation of word representations in vector space. In *Proceedings of ICLR Workshop 2013*, pages 1–12.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013b. Distributed representations of words and phrases and their compositionality. In *Proceedings of NIPS 2013*, pages 1–9.

Tomas Mikolov, Wen tau Yih, and Geoffrey Zweig. 2013c. Linguistic regularities in continuous space word representations. In *Proceedings of NAACL 2013*, pages 746–751.

Toshinobu Ogiso, Mamoru Komachi, Yasuharu Den, and Yuji Matsumoto. 2012. Unidic for early middle Japanese: a dictionary for morphological analysis of classical Japanese. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, pages 911–915.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 2002 Annual Meeting of the Association for Computational Linguistics*, pages 311–318.

Ye Qi, Devendra Sachan, Matthieu Felix, Sarguna Padmanabhan, and Graham Neubig. 2018. When and why are pre-trained word embeddings useful for neural machine translation? In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 529–535.

Hiroyuki Shinnou, Masayuki Asahara, Kanako Komiya, and Minoru Sasaki. 2017. Nwjc2vec: Word embedding data constructed from NINJAL Web Japanese Corpus. *Journal of Natural Language Processing (In Japanese)*, 24(5):705–720.

Daiki Yaginuma, Kanako Komiya, Hiroyuki Shinnou, et al. 2018. WSD of domain adaptation by distributed representation of fine tuning. *IPSJ SIG Technical Reports (NL) (In Japanese)*, 2018(1):1–5.