International Journal of Database Theory and Application
Vol.9, No.5 (2016), pp.121-128
http://dx.doi.org/10.14257/ijdta.2016.9.5.12

Word Sense Disambiguation Based on Perceptron Model

Zhang Chun-Xiang[1, 2], Gao Xue-Yao[3] and Lu Zhi-Mao[4]

[1]*College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China*
[2]*School of Software, Harbin University of Science and Technology, Harbin 150080, China*
[3]*School of Computer Science and Technology, Harbin University of Science and Technology, Harbin 150080, China*
[4]*School of Computer Science and Technology, Dalian University of Technology, Dalian 116024, China*
E-mail: z6c6x666@163.com

Abstract

Word sense disambiguation (*WSD*) *is an important research topic in natural language processing field, which is very useful for machine translation and information retrieval. In this paper, a linear combination model based on multiple discriminative features is proposed to determine correct sense of an ambiguous word, in which morphology and part of speech in left and right words around ambiguous word are used as features. Then, perceptron algorithm is applied to optimize the WSD model. Experiments show that the WSD performance is improved after the proposed method is applied.*

Keywords*: word sense disambiguation; machine translation; information retrieval; linear combination model; multiple discriminative features*

1. Introduction

Brown studies the meaning relatedness at different levels and concludes that a mental lexicon is not composed of separate representations. He finds that word senses have some relations with semantic representation in some degree [1]. Zhong develops an English all-word WSD system with java in which a supervised learning method is adopted. And this system is an extensible and flexible platform for researchers [2]. Stevenson proposes many methods to produce labeled examples automatically for training WSD models. At the same time, a semi-supervised method is adopted to improve WSD's performance in biomedical domain [3]. Li gives a probabilistic model and three instantiations in order to improve the performance of word sense disambiguation. The conditional probability of sense paraphrases is used to determine the best sense in a context. At the same time, a topic model is applied to decompose this conditional probability into two independent ones [4]. Dhillon integrates feature relevance priors of words into WSD model in order to improve the classifier's performance, in which knowledge from similar words is applied to learn priors over discriminative features [5]. Reddy views word sense disambiguation as a distributed constraint optimization problem in which various knowledge sources can be encoded to determine correct senses of ambiguous words [6]. Khapra presents a supervised WSD model with far less annotation knowledge. This method is a middle solution between pure supervised models and pure unsupervised ones. At the same time, the method is not limited to any specific words [7]. Navigli applies sense knowledge in raw text to cluster web search results. A graph-based clustering algorithm is applied to mine triangles and squares in co-occurrence graph of queries and semantic similarity is used to cluster

ISSN: 2005-4270 IJDTA

the search results [8]. Ponzetto gives a novel approach to extend a computational lexicon with encyclopedic relational knowledge. The proposed method is used between WordNet and Wikipedia [9]. Akkaya describes subjectivity word sense disambiguation, which can recognize subjective senses and objective senses automatically in a corpus. Experiments show that the performance of contextual subjectivity and sentiment analysis is improved after subjectivity word sense disambiguation is implemented [10]. Wang proposes a novel approach based on semantic graph structure to determine correct senses of ambiguous words. An undirected weighted semantic graph is used to describe texts in which synsets are viewed as vertices and relationships between synsets are viewed as edges [11]. Quan presents a new WSD method in which few labeled corpus and a large unlabeled corpus are applied to build committee classifiers. Experiments show that the performance of this method is high [12]. Seo uses the relevancy between a target word and raw corpora in word sense disambiguation. At the same time, co-occurrence frequency matrix is utilized to determine senses of many ambiguous words in a sentence [13]. Tufis applies a large parallel corpora to construct a word sense disambiguation classifier, in which word alignment results and word clustering results are utilized. At the same time, 3 sense inventories are adopted to evaluate the performance of the classifier [14]. Rafael presents a semi-supervised method for training WSD models in which unlabeled data are extracted automatically from web and are integrated iteratively into training data set [15].

Morphology and part of speech in left and right words around ambiguous word are extracted as disambiguation features. A linear combination model based on these disambiguation features is constructed for Chinese WSD in this paper. The optimized perceptron classifier is applied to determine correct sense of an ambiguous word. Experimental results show that the disambiguation performance is better.

2. Discriminative Features in WSD Model

Discriminative features and discriminative model are often used to determine correct senses of ambiguous words in word sense disambiguation. Discriminative models are machine learning technology, in which statistical methods are adopted. Discriminative features are often gotten from a given context of an ambiguous word. There are morphology information, part of speech information, sense information and length information in an ambiguous word's context. The information can be mined automatically by natural language processing tools. Their occurrence probabilities are used as discriminative features. Based on these discriminative features, machine learning algorithms are applied in process of word sense disambiguation.

Morphology, part of speech and sense category are three kinds of language knowledge in natural language processing field. The language phenomena that sense categories cover are largest. Part of speech has some ability of covering language phenomena. The language phenomena that morphologies cover are smallest. But, the problem of categorizing senses of different words is very difficult. So, sense category codes are not precise and can not provide more information for word sense disambiguation. In this paper, we extract discriminative features from morphology and part of speech in a given context of an ambiguous word. Here, context of an ambiguous word is the left unit and the right unit around an ambiguous word. Morphology and part of speech in this ambiguous word is not considered. This is because that it is an ambiguous word and there is some uncertainty in it.

The process of extracting morphology and part of speech in a given context is shown in Figure 1.

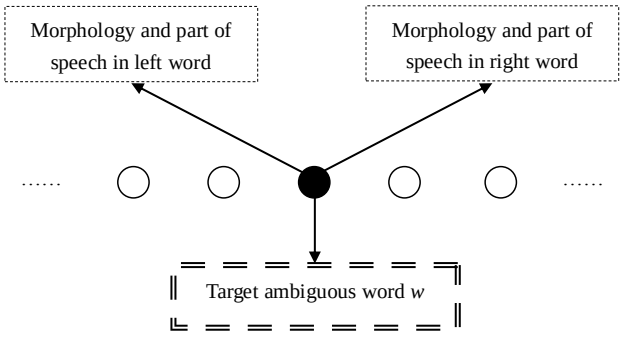

Figure 1. Extract Morphology and Part of Speech in Left and Right Words around Ambiguous Word *W*

3. WSD Classifier Based on Perceptron Algorithm

Perceptron is a supervised learning algorithm for training a binary classifier. Its classification function can decide whether an input belongs to one class or the other one. For Chinese ambiguous word *w*, it has only two semantic categories. The first sense is S_1 and the second sense is S_2. The disambiguation process is viewed as a two classification problem. Perceptron classifier can be used to determine correct sense of ambiguous word *w*. The disambiguation process of ambiguous word *w* is shown in Figure 2. In Figure 2, black dots denote positive instances. The black dot expresses a feature vector of ambiguous word *w* whose sense is S_1 in its context. White dots denote negative instances. The white dot expresses a feature vector of ambiguous word *w* whose sense is S_2 in its context.

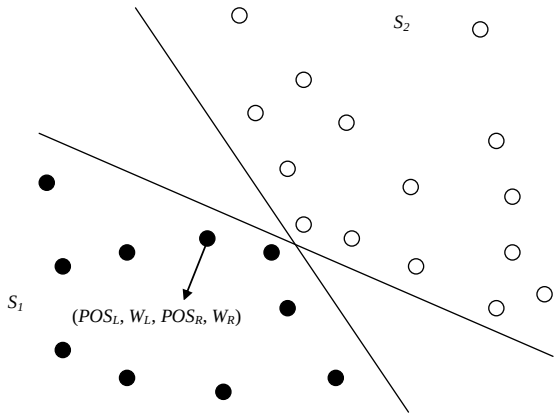

Figure 2. The Disambiguation Process of Ambiguous Word *W*

In Figure 2, a black dot or a white dot denotes feature vector of ambiguous word *w* in its context. The feature vector is described as (POS_L, W_L, POS_R, W_R). There are 4 discriminative features for ambiguous word *w*. They are respectively POS_L, W_L, POS_R and W_R. A linear combination model of multiple discriminative features is used to determine correct sense of ambiguous word *w*. For a given feature vector (POS_L, W_L, POS_R, W_R), its score of $y(POS_L, W_L, POS_R, W_R)$ is calculated as formula (1) describes. If $y(POS_L, W_L, POS_R, W_R)$ is larger, the confidence of it being a positive one is higher. This means that sense of ambiguous word *w* is S_1. If $y(POS_L, W_L, POS_R, W_R)$ is smaller, the confidence of *w* being a negative one is higher. This means that sense of ambiguous word *w* is S_2.

$$y(POS_L, W_L, POS_R, W_R) = \alpha_1 * POS_L + \alpha_2 * W_L + \alpha_3 * POS_R + \alpha_4 * W_R = W \cdot FV \quad (1)$$

Where $W=(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ and $FV=(FV_1, FV_2, FV_3, FV_4)$. Here, $FV_1=POS_L$, $FV_2=W_L$, $FV_3=POS_R$, and $FV_4=W_R$.

This linear combination model is actually a perceptron classifier. So, the classification function can be defined in formula (2).

$$h_{combine}(POS_L, W_L, POS_R, W_R) = y(POS_L, W_L, POS_R, W_R) - \theta = \alpha_1 * POS_L + \alpha_2 * W_L + \alpha_3 * POS_R + \alpha_4 * W_R - \theta \quad (2)$$

The values of parameters α_1, α_2, α_3, α_4 and θ will differ from domains. When $h_{combine}$ is applied to the task of word sense disambiguation in a new domain, we can train parameters automatically to determine these values. Perceptron training algorithm is shown as follows:

Input: Training data set $S=\{(FV_i, y_i)|i=1, 2, \ldots, n\}$, and $y_i \in \{-1, +1\}$ is manual annotation value. FV_i is the feature vector extracted from the ith sentence including ambiguous word w. If the sense of w is S_1, y_i is set to +1. Otherwise, we set -1 to y_i.

Output: values of parameters α_1, α_2, α_3, α_4 and threshold value θ.

(1)Initialization: t=0, $\alpha_i(t)=0.25$, i=1, 2, 3, 4, $\theta=0$.

(2)Do

for each $(FV, y) \in S$

begin

$z = \text{sgn}[\sum_{i=1}^{4} \alpha_i(t) FV_i + \theta(t)]$

if $z \neq y$ then

begin

for i:=1 to 4

$\alpha_i(t+1)=\alpha_i(t)+yFV_i$

$\theta(t+1)=\theta(t)+y$

$t \leftarrow t+1$

end

end

Ideally, we could ask human to annotate sense of ambiguous word w and extract feature vector FV. If the sense of ambiguous word w is S_1, y is set to +1. Otherwise, y is set to -1. Based on these annotated data, we train $h_{combine}(POS_L, W_L, POS_R, W_R)$.

4. Experiments

SemEval-2007 #Task5 is used in order to evaluate the proposed method's performance. This corpus contains 40 Chinese ambiguous words. For each ambiguous word, there are many training data and test data. In training data and test data, every Chinese sentence is segmented into words. Then, human annotators check and modify the segmentation results. Human annotators annotate every word with its part of speech and its sense in English. The format of training corpus and test corpus is shown in Figure 3.

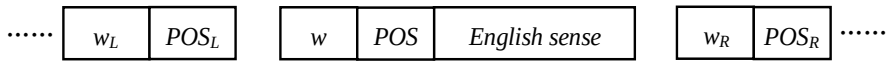

Figure 3. The Format of Training Corpus and Test Corpus

In Figure 3, there are two kinds of language knowledge for every word unit. They are respectively morphology and part of speech. These language knowledge can help WSD model to determine correct sense of ambiguous word *w*. For ambiguous word *w*, its English sense is provided. For ambiguous word 'Zhong Yi', its left word unit, its word unit and its right word unit are shown in Figure 4. In Figure 4, v denotes a verb and n expresses a noun. Its left word is 'Zuan Yan' and its part of speech is v. Its right word is 'Li Lun' and its part of speech is n. The ambiguous word is 'Zhong Yi' and its part of speech is n. The sense in English is described as traditional Chinese medical science.

… Zuan Yan/v Zhong Yi/n/traditional_Chinese_medical_science Li Lun/n…

Figure 4. The Example of Annotation Corpus

We select four common ambiguous words including 'Zhong Yi', 'Cai', 'Tui Fan' and 'Biao Mian'. For every ambiguous word, its training corpus is used to train WSD model and its test corpus is applied to evaluate the performance of the optimized classifier.

In order to evaluate the proposed method's performance, two experiments are designed. In experiment 1, left word and right word around ambiguous word are extracted as disambiguation features. Based on these disambiguation features, a bayesian model is adopted to construct a classifier to determine correct sense of this ambiguous word. Training corpus including ambiguous word *w* is utilized to train this ambiguous word's WSD classifier. This classifier is only used to determine the sense of ambiguous word *w*. In experiment 2, formula (2) is adopted to build WSD classifier, and perceptron training algorithm is employed to get optimized parameters α_1, α_2, α_3, α_4 and θ.

For every ambiguous word, an optimized WSD classifier is gotten. The purpose is to decrease the difficulty of constructing WSD classifier. This is because that there is some influence between two different ambiguous words on word sense disambiguation. This method can improve the disambiguation performance.

The accuracy rate of disambiguation is illustrated in Table 1. Table 1 shows that accuracy rate in experiment 2 is higher than or equal to that of experiment 1. For word 'Tui Fan', the improvement of accuracy rate is highest and it achieves 30%. For word 'Cai, its accuracy rate of disambiguation keeps unchanged. For word 'Zhong Yi', the improvement of accuracy rate is 25%. For word 'Biao Mian', the improvement of accuracy rate is 5.6%.

Table 1. The Accuracy Rate of Disambiguation

Ambiguous words	Experiment 1	Experiment 2
Zhong Yi	37.5%	62.5%
Cai	33.3%	33.3%
Tui Fan	50.0%	80.0%
Biao Mian	50.0%	55.6%

The proposed WSD model has two advantages. The first one is that morphology and part of speech are utilized as discriminative features. More language knowledge is adopted in process of determining correct senses of ambiguous words. So, the model has more discriminative ability. The second one is that perception model is applied in WSD process. The perception model has more discriminative ability than bayesian model.

5. Conclusion

In this paper, language knowledge in left and right word units around ambiguous word is applied in word sense disambiguation. In order to utilize language knowledge fully, a linear combination model based on multiple discriminative features is constructed to determine correct senses of ambiguous words. Perceptron algorithm is used to train the linear combination model. Experimental results show that accuracy rate of WSD is improved by this method.

Acknowledgements

This work is supported by China Postdoctoral Science Foundation Funded Project(2014M560249) and Natural Science Foundation of Heilongjiang Province of China(F2015041).

References

[1] S. W. Brown, "Choosing sense distinctions for WSD: psycholinguistic evidence", Proceedings of ACL-08: HLT, **(2008)**, pp. 249-252.
[2] Z. Zhong and H. T. Ng, "It makes sense: a wide-coverage word sense disambiguation system for free text", Proceedings of the ACL 2010 System Demonstrations, **(2010)**, pp. 78-83.
[3] M. Stevenson and Y. K. Guo, "The effect of ambiguity on the automated acquisition of WSD examples", Human Language Technologies: the 2010 Annual Conference of the North American Chapter of the ACL, **(2010)**, pp. 353-356.
[4] L. L. Li, B. Roth and C. Sporleder, "Topic models for word sense disambiguation and token-based idiom detection", Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, **(2010)**, pp. 1138-1147.
[5] P. S. Dhillon and L. H. Ungar, "Transfer learning, feature selection and word sense disambiguation", Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, **(2009)**, pp. 257-260.
[6] S. Reddy and A. Inumella, "WSD as a distributed constraint optimization problem", Proceedings of the ACL 2010 Student Research Workshop, **(2010)**, pp. 13-18.
[7] M. M. Khapra, A. Kulkarni and S. Sohoney, "All words domain adapted WSD: finding a middle ground between supervision and unsupervision", Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, **(2010)**, pp. 1532-1541.
[8] R. Navigli and G. Crisafulli, "Inducing word senses to improve web search result clustering", Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, **(2010)**, pp. 116-126.
[9] S. P. Ponzetto and R. Navigli, "Knowledge-rich word sense disambiguation rivaling supervised systems", Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, **(2010)**, pp. 1522-1531.
[10] C. Akkaya, J. Wiebe and R. Mihalcea, "Subjectivity word sense disambiguation", Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, **(2009)**, pp. 190-199.
[11] J. H. Wang, J. Y. Liu and C. Wang, "Word sense disambiguation with semantic graph structure", Journal of Beijing University of Posts and Telecommunications, vol. 29, no. 2, **(2006)**, pp. 96-100.
[12] C. Q. Quan, T. T. He and D. H. Ji, "Word sense disambiguation based on multi-classifier decision", Computer Research and Development, vol. 43, no. 5, **(2006)**, pp. 933-939.
[13] H. C. Seo, H. C. Rim and M. G. Jang, "Word sense disambiguation by relative selection", Proceedings of the 2nd International Joint Conference on Natural Language Processing, **(2005)**, pp. 920-932.
[14] D. Tufis, "Word sense disambiguation: a case study on the granularity of sense distinctions", WSEAS Transactions on Information Science and Applications, vol. 2, no. 2, **(2005)**, pp. 183-188.
[15] G. C. Rafael and P. Rosso, "Semi-supervised word sense disambiguation using the web as corpus", Proceedings of the 10th International Conference on Computational Linguistics and Intelligent Text Processing, **(2009)**, pp. 256-265.

Authors

Chun-Xiang Zhang, is Ph.D. and graduates from Ministry of Education-Microsoft Key Laboratory of Natural Language Processing and Speech, School of Computer Science and Technology, in Harbin Institute of Technology. He is also a professor in Harbin University of Science and Technology. His research interests are natural language processing, machine translation and machine learning. He has authored and coauthored more than fifty journal and conference papers in these areas.