International Journal of Database and Theory and Application
Vol.9, No.2 (2016), pp.17-22
http://dx.doi.org/10.14257/ijdta.2016.9.2.03

Determine Word Sense Based on Semantic and Syntax Information

Zhang Chun-Xiang[1, 2], Sun Lu-Rong[3] and Gao Xue-Yao[3*]

[1]*School of Software, Harbin University of Science and Technology, Harbin 150080, China*

[2]*College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China*

[3]*School of Computer Science and Technology, Harbin University of Science and Technology, Harbin 150080, China*

**Corresponding author E-mail: gaoxueyao@hrbust.edu.cn (*Gao Xue-Yao*)*

Abstract

Word sense disambiguation (WSD) plays an important role in natural language processing fields. Semantic category is semantic knowledge and part-of-speech is syntax knowledge. In this paper, word window is opened to get semantic category and part-of-speech of left and right adjacent words around an ambiguous word. A new approach of determining true meanings of ambiguous words based on support vector machine (SVM) is given. The training corpus in SemEval-2007: Task#5 is applied to optimize SVM and the optimized SVM is tested. Experimental results show that the performance of the proposed method is improved.

***Keywords**: word sense disambiguation; natural language processing; semantic knowledge; part-of-speech; support vector machine*

1. Introduction

Ambiguous words always have an influence on the development of information processing. The task of word sense disambiguation is to determine correct sense of an ambiguous word in a specific context. Word sense disambiguation is mainly divided into three kinds of approaches. They are respectively supervised approach, unsupervised method and semi-supervision algorithm. Supervised approach is effective for word sense disambiguation. But this approach needs a lot of human annotated corpus. Unsupervised method can solve the problem of knowledge acquisition. But the disambiguation result is poor. In semi-supervised algorithm, labeled corpora and unlabeled corpora are used to train a classifier.

David presents a supervised method based on decision lists, which minimizes the training data fragmentation for decision trees [1]. Wu presents a feature-based word sense disambiguation method and distributional features from grammatical knowledge base of contemporary Chinese are used. At the same time, bayesian model is used as WSD classifier [2]. Ellman exploits and extends information in the Hector dictionary. Several filters are applied to prune the candidate set of word senses and the most frequent one is selected [3]. Popescu studies the importance of the clustering method in unsupervised word sense disambiguation. He finds that a perfect clustering algorithm can make up for the shortage of external knowledge [4]. Karov uses a text corpus and a machine readable dictionary to determine senses of ambiguous words, in which word similarity and context similarity measures are applied [5]. Bordes gives a neural network architecture to keep and enhance original data. The purpose is to integrate multi-relational graphs into a flexible continuous vector space [6]. Special presents a new WSD method in which an

ISSN: 1975-0080 IJDTA

inductive logic programming method is applied to get theories from first-order logic representations. The corpus-based evidence can be combined closely with deep and shallow background knowledge [7]. Navigli proposes a graph-based model with few parameters. The process of training model does not need sense-annotated data. Experiments show that the lexicon and the graph connectivity influence WSD performance [8]. Wang uses a pattern-based method to get pseudo samples. Pseudo samples and sense tagged data are combined to estimate conditional probabilities of variables. The purpose is to improve the performance of WSD classifier [9]. Schwab studies the influence of 3 unsupervised algorithms on word sense disambiguation, including genetic algorithm, simulated annealing algorithm and ant colony algorithm. Experiment results show that the performance of ant colony algorithm is better and a shorter time is needed [10]. Imamura combines active learning method with semi-supervised learning method to train WSD classifier, in which pseudo negative examples and unlabeled examples are used [11]. Liu presents a new unsupervised word sense disambiguation algorithm based on Google distance. Information distance and Kolmogorov complexity are applied to compute the Google distances between words and phrases [12]. Huang assigns positional weights to contextual words and computes the context similarity between a new instance and pre-labeled instances. Then, the context similarity is applied to word sense disambiguation [13]. Agirre describes a new disambiguation model in which random walk algorithm and large lexical knowledge bases are used. Experiments show that the new method exceeds other graph-based ones in precision [14].

In this paper, we use semantic information and part-of-speech as discriminative features. Support vector machine is applied to determine correct meanings of ambiguous words. Experimental results show that the precision of disambiguation is improved.

2. Extracting Part-of-Speech and Semantic Information

Disambiguation features are extracted from contexts of ambiguous words. Semantic contexts can be chapters, paragraphs, sentences and windows containing ambiguous words. Disambiguation features may be word form, part-of-speech, semantic category and length. In this paper, word window containing ambiguous word is regarded as semantic context. Semantic category and part-of-speech are extracted as discriminative features. Semantic categories of left and right adjacent words are gotten from Tongyici Cilin. Tongyici Cilin is a Chinese semantic lexicon and provides semantic categories for words. Semantic category of a word has three layers. The first one is big category, the second one is middle category and the last one is small category. For Chinese word ‘zhong yi’, its sense category is Dk03. Its big category is D, middle category is k and small category is 03. Here, we only use big category as discriminative feature to reduce the effects of data sparseness. For Chinese sentence containing ‘zhong yi’, the process of extracting disambiguation features is shown in Figure 1.

Firstly, Chinese sentence which contains ambiguous word is segmented into words. The result is ‘ta/ nu li/ fa jue/ chuan tong/ zhong yi/ de/ jing cui/ ./’. Secondly, we use part-of-speech tagging tool to label every word automatically in this sentence and the result is ‘ta/r nu li/ad fa jue/v chuan tong/a zhong yi/ng de/u jing cui/ng ./w’. In Chinese sentence, target ambiguous word is searched for disambiguation. Word ‘zhong yi’ is viewed as the center and a word window is opened to obtain discriminative information. For target word, its left adjacent word is ‘chuan tong/a’ and its right adjacent word is ‘de/u’. They are extracted from word window. Then, part-of-speech of left and right adjacent words can be obtained. Part-of-speech of ‘chuan tong’ is a, and part-of-speech of ‘de’ is u. Thirdly, semantic codes of left word and right word are gotten from Tongyici Cilin.

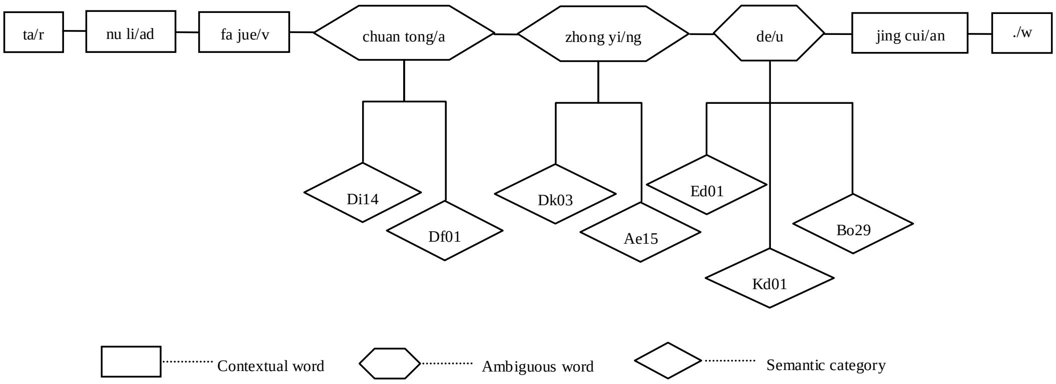

Figure 1. Extract Disambiguation Features

When word window is extended, it contains more discriminative information. In fact, it is difficult to obtain a lot of human annotated corpus. If the window is opened too widely, it will lead to data sparseness. So, disambiguation features are extracted from left and right adjacent word units around an ambiguous word.

Disambiguation features are respectively POS_{Left}, Sem_{Left}, POS_{Right} and Sem_{Right}. POS_{Left} is its left word's part-of-speech. Sem_{Left} represents semantic category of the left word. POS_{Right} is its right word's part-of-speech. Sem_{Right} denotes semantic category of the right word. In the above sentence, 'de' is also an ambiguous word and cannot provide any guidance for the disambiguation process. Dice coefficient is a measurement function of similar degree. In this paper, dice coefficient is used to determine semantic category of an ambiguous word in a specific context. Then, discriminative features of ambiguous word 'zhong yi' are gotten. They are respectively a, D, u, K. Here, w is an ambiguous word and has m kinds of different senses. They are respectively s_1, s_2, ..., s_m.

Probabilities of disambiguation features are estimated on training data. Probabilities of POS_{Left} and POS_{Right} are shown in equation (1).

$$P(POS_i) = \frac{Number(POS_i)}{n} \qquad i = Left, Right \tag{1}$$

Probabilities of Sem_{Left} and Sem_{Right} are computed in equation (2).

$$P(Sem_i) = \frac{Number(Sem_i)}{n} \qquad i = Left, Right \tag{2}$$

Here, $Number(X)$ is the number of X and n is the number of sentences in training data.

3. Word Sense Disambiguation Classifier Based on SVM

Support vector machine is a statistical learning model proposed by Vapnik for linear classification problems. For binary classification problems, SVM is used to construct a hyperplane in feature space to maximize interval between positive examples and negative ones. We use SVM classifier to select an optimal hyperplane $w^* \cdot x + \theta^* = 0$, in which feature space is divided into two different categories.

In word sense disambiguation, corpus annotated with semantic categories is linearly inseparable. Nonlinear support vector machine is chosen for word sense disambiguation. Kernel function is used in SVM model. Disambiguation features are mapped into a high dimensional space. Then, the linearly inseparable problem is changed to a linear separable one. Semantic category features (P(Sem_{Left}), P(Sem_{Right})) and part-of-speech features (P(POS_{Left}), P(POS_{Right})) are utilized to construct a word sense disambiguation classifier.

In order to improve the performance of disambiguation classifier, human annotated corpus is applied to train parameters w and θ. The solving process will be turned into a quadratic programming problem[15].

$$\min_{\alpha} \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^{n}\alpha_i$$

$$s.t. \quad \sum_{i=1}^{n}\alpha_i y_i = 0 \qquad 0 \le \alpha_i \le C \qquad i = 1,2,...,n$$

$K(x_i, x_j)$ is kernel function. Here, $x_i = (P(Sem^i_{Left}), P(POS^i_{Left}), P(Sem^i_{Right}), P(POS^i_{Right}))$, $y_i = \{-1,+1\}$, i=1, 2, …, n.

The optimal solution is shown as follows:

$$\alpha^* = (\alpha_1^*, \alpha_2^*, ..., \alpha_n^*)$$

Several sentences containing w are selected. These sentences in which semantic category of w is annotated with s_j are viewed as positive examples. Other sentences are negative instances. The binary classification SVM is used to construct decision function $f_j(x)$ shown in equation (3).

$$f_j(x) = \sum_{i=1}^{n}\alpha_i^j y_i K(x_i, x) + \theta_j \tag{3}$$

The process of calculating parameter θ_j is shown as follows:

$$\theta_j = y_j - \sum_{i=1}^{n}\alpha_i^j y_i K(x_i, x_j)$$

For ambiguous word w, the disambiguation classifier is designed in equation (4).

$$s = \arg\max_{j=1,2,...,m} f_j(x) \tag{4}$$

4. Experiments

SemEval-2007: Task#5 is applied as training corpus and test corpus. 5 common ambiguous words are selected to measure the performance of this method. They are respectively ‘ri zi’, ‘dui wu’, ‘gan’, ‘tian di’ and ‘tiao’. The distribution of training corpus and test corpus is shown in Table 1.

In order to prove this method’s validity, two groups of comparative experiments are conducted. A word window is opened to extract disambiguation features, including left and right adjacent words of an ambiguous word. In experiment 1, a classifier based on bayes model is trained in which left and right adjacent words are used as discriminative features. The optimized bayes model is applied to determine senses of ambiguous words in test corpus. In experiment 2, a classifier based on SVM model is built in which semantic category and part-of-speech are used as disambiguation features. The optimized SVM model is applied to determine senses of ambiguous words in test corpus.

Table 1. The Distribution of Training Corpus and Test Corpus

Ambiguous word	The number of sentences in training corpus	The number of sentences in test corpus	The total number of sentences
‘ri zi’	88	32	120
‘dui wu’	64	22	86
‘gan’	56	18	74
‘tian di’	65	25	90
‘tiao’	40	14	54

Automatic semantic annotation is compared with manual semantic one. Then, the disambiguation accuracy of test corpus is calculated. The results are shown in Table 2.

Table 2. The Disambiguation Accuracy of Test Corpus

Ambiguous words	Experiment 1	Experiment 2
'ri zi'	46.88%	78.13%
'dui wu'	36.36%	81.82%
'gan'	27.78%	55.56%
'tian di'	72.00%	76.00%
'tiao'	50.00%	71.43%

From Table 2, it can be seen that accuracy rate in experiment 2 is better than that in experiment 1. For word 'ri zi', 'dui wu', 'gan', 'tian di' and 'tiao', the disambiguation accuracy is improved. Among them, accuracy rate of 'dui wu' increases 45.46%. The disambiguation accuracy of 'tian di' increases 4.00%.

In experiment 1, word form provides less guidance information for WSD. In experiment 2, part-of-speech has some ability of covering language phenomena, and semantic categories can cover more language phenomena. They provide more knowledge for word sense disambiguation and improve semantic identification ability of the model. In addition, semantic category and part-of-speech also have some generalization ability. When model parameters are estimated, data sparseness can be partly avoided in experiment 2.

From experimental results, we can see that when semantic category and part-of-speech are viewed as discriminative features and SVM is used as WSD classifier, the disambiguation accuracy is increased.

5. Conclusion

In this paper, an ambiguous word is viewed as the center. Semantic category and part-of-speech from left and right adjacent words are applied to WSD. In training corpus, parameters of disambiguation features are estimated. Support vector machine is used to determine correct meanings of ambiguous words. Comparative experiments are conducted and experimental results show that after the new method is applied, the accuracy of disambiguation is improved.

Acknowledgement

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

References

[1] Y. David, "Hierarchical decision lists for word sense disambiguation", Computers and the Humanities, vol. 34, no. 1, **(2000)**, pp. 179-186.
[2] Y. F. Wu, P. Jin and T. Guo, "Coarse-grained word sense disambiguation using features described in the lexicon", Journal of Chinese Information Processing, vol. 21, no. 2, **(2007)**, pp. 3-8.
[3] J. Ellman, I. Klincke and J. Tait, "Word sense disambiguation by information filtering and extraction", Computers and the Humanities, vol. 34, no. 1, **(2000)**, pp. 127-134.
[4] M. Popescu and F. Hristea, "State of the art versus classical clustering for unsupervised word sense disambiguation", Artificial Intelligence Review, vol. 35, no. 3, **(2011)**, pp. 241-264.
[5] Y. Karov and S. Edelman, "Similarity-based word sense disambiguation", Computational Linguistics, vol. 24, no. 1, **(1998)**, pp. 41-59.
[6] A. Bordes, X. Glorot, J. Weston and Y. Bengio, "A semantic matching energy function for learning with multi-relational data: Application to word-sense disambiguation", Machine Learning, vol. 94, no. 2, **(2014)**, pp. 233-259.

[7] L. Specia, M. Stevenson and M. D. G. V. Nunes, “Assessing the contribution of shallow and deep knowledge sources for word sense disambiguation”, Language Resources and Evaluation, vol. 44, no. 4, **(2010)**, pp. 295-313.
[8] R. Navigli and M. Lapata, “An experimental study of graph connectivity for unsupervised word sense disambiguation”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 4, **(2009)**, pp. 678-692.
[9] X. Wang and Y. Matsumoto, “Improving word sense disambiguation by pseudo-samples”, First International Joint Conference on Natural Language Processing, **(2004)**, pp. 386-395.
[10] D. Schwab, J. Goulian and A. Tchechmedjiev, “Worst-case complexity and empirical evaluation of artificial intelligence methods for unsupervised word sense disambiguation”, International Journal of Web Engineering and Technology, vol. 8, no. 2, **(2013)**, pp. 124-153.
[11] M. Imamura, Y. Takayama and N. Kaji, “A combination of active learning and semi-supervised learning starting with positive and unlabeled examples for word sense disambiguation: an empirical study on Japanese web search query”, Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, **(2009)**, pp. 61-64.
[12] P. Y. Liu, Y. Z. Xue, S. Q. Li and S. Liu, “Minimum normalized Google distance for unsupervised multilingual Chinese-English word sense disambiguation”, Proceedings of the 4th International Conference on Genetic and Evolutionary Computing, **(2010)**, pp. 252-255.
[13] S. Huang, X. Zheng, H. Kang and D. Chen, “Word sense disambiguation based on positional weighted context”, Journal of Information Science, vol. 38, no. 2, **(2013)**, pp. 225-237.
[14] E. Agirre, O. L. D. Lacalle and A. Soroa, “Random walks for knowledge-based word sense disambiguation”, Computational Linguistics, vol. 40, no. 1, **(2014)**, pp. 57-84.
[15] Y. Yao, H. Zhao and Z. J. Liu, “Multi-class classifier of non-linear SVM decision tree”, Journal of Xidian University, vol. 34, no. 6, **(2007)**, pp. 873-876.

Author

Chun-Xiang Zhang, He 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. At the same time, he is a post doctor in College of Information and Communication Engineering, in Harbin Engineering University. His research interests are natural language processing and machine learning. He has authored and coauthored more than fifty journal and conference papers in these areas.

