Classification of Chinese Word Semantic Relations

Changliang Li[1], Teng Ma[1, 2]

[1] Institute of Automation, Chinese Academy of Sciences, Beijing, P.R. China
changliang.li@ia.ac.cn
[2] School of Mathematics and Statistics, Wuhan University, Wuhan Hubei 430072, China
mteng@whu.edu.cn

Abstract. Classification of word semantic relation is a challenging task in natural language processing (NLP) field. In many practical applications, we need to distinguish words with different semantic relations. Much work relies on semantic resources such as Tongyici Cilin and HowNet, which are limited by the quality and size. Recently, methods based on word embedding have received increasing attention for their flexibility and effectiveness in many NLP tasks. Furthermore, word vector offset implies words semantic relation to some extent. This paper proposes a novel framework for identifying the Chinese word semantic relation. We combine semantic dictionary, word vector and linguistic knowledge into a classification system. We conduct experiments on the Chinese Word Semantic Relation Classification shared task of NLPCC 2017. We rank No.1 with the result of F1 value 0.859. The results demonstrate that our method is very scientific and effective.

Keywords: Word relation classification, Word vector, Semantic lexicons, Linguistic knowledge.

1 Introduction

The classification of word semantic relation focuses on lexical level, which purpose is to predict the categorization of semantic relation between two Chinese words. Specifically, given a pair of Chinese words, it is required to recognize this pair of words into one of the following semantic relations: synonym, antonym, hyponym and meronym. In many applications, we need to distinguish words with different semantic meanings, such as information extraction and the construction of semantic networks [1, 2]. This task is a challenging work in natural language processing field.

Some attempts focus on dealing with this problem. One way is to rely on manual semantic resources, such as Tongyici Cilin [3] and HowNet [4]. The semantic relation of some words has been labelled in these resources. It is effective and simple to find the semantic relation by matching dictionaries. However, the drawback is also obvious that we can only recognize the pairs when both of members are presented in the lexicons.

The traditional classification methods utilize various features to train a classifier in a supervised manner. And these features can include lexical bag-of-words features and other features based on syntactic parse trees. For grammar parsing trees, the path of dependencies between the constituency and the target entity has proven to be useful [5, 6].

In recent years, researchers pay attention to word embedding methods, such as skip-gram model [7-9] and Glove model [10]. This kind of methods has demonstrated outstanding performance in various tasks. Word embeddings are supposed to capture useful syntactic and semantic properties [11]. However, basic embedding methods have drawbacks in nature. One of the limitations is that word embeddings are usually learned by predicting a target word with its local context, leading to only limit information being captured. Therefore, researchers raise interest in integrating lexicons into word embeddings to capture multiple semantics [12-16].

The introduction of the word embeddings provides a new point for our task. Word embeddings are trained by predicting words between noun pairs using lexical relation-specific features on a large unlabeled corpus, which absorbed incorporate relation-specific information into the word embeddings [17]. [18] founded that the learned word representations capture meaningful syntactic and semantic regularities in a very simple way: vector offset.

In this paper, we propose a novel framework that combines word vectors, semantic lexicon and linguistic knowledge for Chinese word semantic relation classification. We firstly match a pair of words with semantic dictionary, and then our system will output the label of these words if matched successful. Otherwise, we will automatically recognize the semantic relationship of these words with supervised model. In this model, we extract three features and combine them as our model input. Besides, we also introduce linguistic knowledge to our model. We conduct experiment on Chinese Word Semantic Classification shared task [19], which provides a benchmark dataset for evaluating the study on word relation classification for Chinese language. And we rank No.1 with the result of 0.859 of F1 value and outperform other methods by a very large margin.

The remainder of this paper is arranged as follows. In section 2, we introduce our method. Section 3 introduces our dataset and experiment settings. Section 4 presents the experiments. Some conclusions are summarized in Section 5.

2 Methodology

In this section, we introduce our classification system firstly and then explain each part of the system.

2.1 Classification System

The method consists of three aspects, semantic dictionaries, supervised model and linguistic knowledge. We match a pair of words with a semantic dictionary. If the match is successful, we will output the label of the words in the dictionary. Otherwise, we will utilize supervised model to classify the words semantic relation automatically. Figure 1 gives the overall illustration of our method.

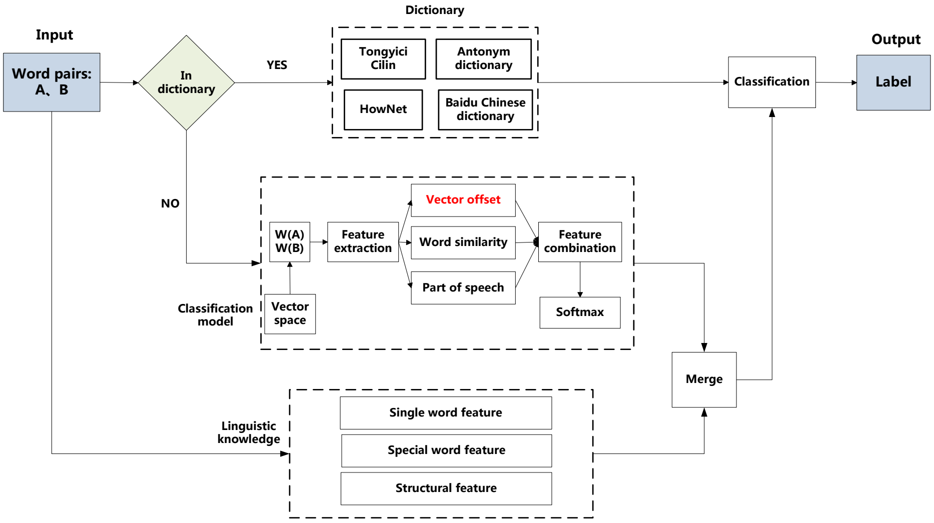

Fig. 1. The overall illustration of our method

2.2 Semantic Relation Classification Based on Dictionary

Tongyici Cilin [3]**.** This dictionary has a lot of applications. One of which is used to find synonyms of words. So we can use the dictionary to recognize the semantic relationship of a pair of words, and the relationship will be labelled synonyms if these words both involved in the dictionary.

The Cilin dictionary is organized by a 5-layer hierarchical structure. Correspondingly, it supplies 5-layer patterns to generate code for a group of words. And there are three ways to end the code items in the dictionary: "=", "#", "@". "=" stands for synonyms, "#" stands for related words, but not synonymous, "@" represents isolated words, neither synonyms nor related words. For instance, the pair of words (水平/level, 程度/degree) are coded as "Dd12A02=", words (男队/male team, 女队/female team) are coded as "Dd07B09#", and the word (高地/highland) is coded as "Dd09A05@". So we only need those words which code items ended by "=".

HowNet [4]**.** HowNet is a common knowledge base, which is based on the concepts represented by Chinese and English terms as a description object to reveal the relationship between concepts and concepts and attributes possessed by concepts. HowNet also reflects the semantic relations among words, including hyponyms, synonyms, antonyms and meronyms and so on. So we can use HowNet to find the semantic relationship of two words.

Antonym Dictionary. There are many antonym phrases on the Internet, so we can construct an antonym dictionary which contains high-frequency antonyms. We match a pair of words with the antonyms dictionary, and if the match is successful, then the semantic relationship will be labelled antonyms.

Other Dictionary. We also use the Internet to recognize the semantic relation of a pair of words. Synonyms and antonyms of a word are listed in the Baidu Chinese Dictionary. Through this resource, we can automatically match a pair of words. For example, there is a pair of words (高兴/happy, 难过/sad), if we enter the website: http://hanyu.baidu.com/s?wd=高兴&from=zici, then we will find that the relationship of two words is antonyms. Figure 2 shows the example, and both words are labelled in red blocks.

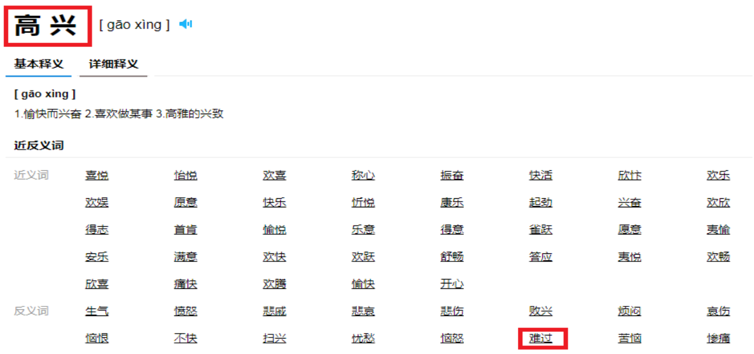

Fig. 2. The example of Baidu Chinese dictionary

2.3 Semantic Relation Classification Based on Various Features

In this section, we explain the classification model based on various features: vector offset, word similarity and part of speech. This method mainly uses the word embedding to extract the features, and recognize the semantic relation of two words. We will extract features from different angles.

Vector Offset. Vector offset technique is employed in this step, which is a simple algebraic operation performed on the word vectors. For example, vector(King) – vector(Man) + vector(Woman) produces a vector that is closest to the vector representation of the word Queen [18]. As a result, the male/female relationship is learned automatically.

We approximate the above example as a mathematical expression: vector (King) – vector (Queen) ≈ vector (Man) – vector (Woman). We can get some information from this formula. We know that King and Queen are very similar words. This pair of words has many same attributes, such as noble identity, style of the palace, luxury and so on. A significant difference between these two words is that the gender is different, King is male, and Queen is female. Similarly, the main difference between Man and

Woman is the gender either. From this point of view, the above formula can depict different semantic parts of two words.

In this way, the model can capture the semantic regularities between a pair of words easily. So we get the first feature of the model.

Given a pair of words (v_{s_1}, v_{s_2}), the feature defines v_{info} as follows:

$$v_{info} = v_{s\_1} - v_{s\_2} \tag{1}$$

Where v_{s_1} and v_{s_2} are vector representations of word w_{s_1} and w_{s_2} .

Cosine Similarity. We extract new feature to characterize semantic similarity between two words. In the field of machine learning, especially in the field of NLP, cosine similarity is often used to calculate the text similarity. So here we depict the semantic similarity by calculating the cosine value of two words vectors. Given a word pair v_{s_1}, v_{s_2}, we use c^* to represent cosine similarity.

Part of Speech. In Chinese words, the words' part of speech can give us a lot of information. If two words are both nouns, their semantic relationship is probably not antonymous, because antonyms are often adjectives or verbs. Under this assumption, we extract part of speech as a feature, and for simplicity, we divide the feature into four parts: nouns, verbs, adjectives, and others, and then encode them. Given a word w_{s_1}, the feature defines p_{s_1} as follows:

$$p_{s\_1} = \begin{cases} 1, & if\ w_{s\_1}\ is\ adjective. \\ 2, & if\ w_{s\_1}\ is\ verb. \\ 3, & if\ w_{s\_1}\ is\ noun. \\ 4, & otherwise. \end{cases} \tag{2}$$

Classification Model. Now we have obtained three features of depicting semantic relations: v_{info}, c^*, p_{s_i} . p_{s_i} stands for the part of speech of the word w_{s_i}. We combine these features and feed it as our model input.

Given a word pair v_{s_1}, v_{s_2}, we combine these features into a vector defines X:

$$\mathrm{X} = (v_{info}, c^*, p_{s\_1}, p_{s\_2}) \tag{3}$$

Softmax classifier is a supervised model for multi-classification problems, which is very simple but effective. Here, we use the softmax classifier to recognize our data.

2.4 Semantic Relation Classification Based on Linguistic Knowledge

Chinese words have many linguistic features. We can apply these characteristics to our task so as to improve the accuracy of our model. In the previous section, we use the softmax classification model to recognize the semantic relation of words. In this section, we will introduce the characteristics of Chinese word to update the output of the model.

Structural Feature. In Chinese vocabulary, some words have special word structure, and we can use the structure to update the results of our model. For instance, the semantic relation of “树”(tree) and “树枝”(branch), “花”(flower) and “花蕾”(bud), “龟”(turtle) and “龟壳”(turtle shell) are all meronyms. More importantly, these words have the same structure. Their structure is “A and AB”. That phenomenon is very common to Chinese vocabulary, so we can use this structural feature into our system. Here is one more example, the semantic relation of “花”(flower) and “玫瑰花”(rose), “蛇”(snake) and “海蛇”(sea snake), “水”(water) and “海水”(sea water) are all hyponyms. And their word structure is “A(蛇) and BA(海蛇)” or “A(花) and BCA(玫瑰花)”, anyway, “A” is in the end. In addition, there are many other structural features can be for our reference, and we can make use of the structural features to improve the accuracy of our system.

Special word feature. In Chinese vocabulary, some words have no practical meaning, but they are part of a word, such as “子” in the word “椅子”. This type of word structure can also help our task. For example, the semantic relationship of “桌子” and “桌腿” is meronyms, and if we remove the “子” from the “桌子”, then we will find we can use the word structural feature to update the results of our model.

Single Word Feature. There are many words in the Chinese vocabulary with single word, such as “大”(big) and “小”(small), “胖”(fat) and “瘦”(thin). And normally, these words are antonyms. This is not an accidental phenomenon. In Chinese vocabulary, few synonyms, hyponyms, or meronyms are composed of single word, they are often composed of two words or more words, such as “开心”(happy) and “高兴”(happy), “车票”(ticket) and “火车票”(train ticket), “钢笔”(pen) and “笔帽”(pen cap). So if both words are composed of a single word, then we can make sure that the semantic relationship of the pair of words is antonyms.

3 Experiment Settings

3.1 Data Set and Evaluation

The proposed approach is evaluated on the dataset released by NLPCC2017 shared task 1 [19]. The dataset contains 200 sample data and 2000 test word pairs with their semantic category. Our training set contains 913-word pairs, which come from web search and manual annotation, which consist of synonyms, antonyms, hyponyms and meronyms. We trained supervised model with training dataset and used the 200 sample data as validation set.

The performance of our experiments is evaluated by the macro-averaged precision (P), recall (R) and F1-score. So we compute P, R, and F1-score for each relation, and then compute the macro-averaged P, R and F1-score.

4 Results and Analysis

In this section, we show the results of our experiments and analyze the results.

Table 1 shows the results of our entire classification system, including the synonym class, the antonym class, the hyponym class, the meronym class and the macro-averaged score. In Table 1, P represents the precision, R stands for the recall. Our scoring metric is macro-averaged F1-score for four-way classification.

Table 1. Result of our classification system

Category	P	R	F1
Synonym	0.859	0.952	0.903
Antonym	0.945	0.828	0.883
Hyponym	0.770	0.870	0.817
Meronym	0.885	0.784	0.831
Macro-average	**0.865**	**0.859**	**0.859**

As is shown in Table 1, four categories of F1 value have exceeded 0.8, and F1 value of synonym reached 0.903. Specifically, the score of synonym and antonym is significantly better than hyponym and meronym, and their F1-score are both higher than 0.85, indicating that our method is very effective for synonym and antonym. As for the hyponym and meronym, the hyponym P value and the meronym R value are slightly lower. It is due to that in Chinese vocabulary, the semantic relationship of some words is difficult to recognize in these two categories. For example, the semantic relationship of “四大名著”(Four famous novels) and “红楼梦”(Story Stone) is not only hyponyms but also meronyms. In addition, our training data set is relatively small, which will restrict the result of our classification system.

Table 2 shows the comparison of the supervised model and the merging method. In Table 2, SMX means softmax method; DC means dictionary method; LK means linguistic knowledge method. We can see that the result of ID.1 reaches 0.527, which is acceptable given the small training dataset. Besides, the result ID.3 achieves 0.859 of F1 value, which performs 0.332 (63%) higher than ID.1. It illustrates the effectiveness of the merging approach.

Table 2. Results of classification by various models or methods

ID.	Method	P	R	F1
1	SMX	0.576	0.526	0.527
2	SMX + DC	0.761	0.736	0.729
3	SMX + DC + LK	**0.865**	**0.859**	**0.859**

Table 3 shows the results of the classification system by selecting different features. It can be seen that F1 value of ID.1 reaches 0.790, and the result of all features is 0.859, indicating that the feature v_{info} is critical to the whole classification system. Besides, the F1 value of ID.2 performs 0.036 (4.6%) higher than ID.1 duo to the use of feature c^*. The result of ID.3 reaches 0.859 by introducing feature p_{s_i} , and ID.3 is submitted as our final result.

Table 3. Comparison between single and merging features

ID.	Strategy	P	R	F1
1	v_{info}	0.825	0.789	0.790
2	$v_{info} + c^*$	0.843	0.827	0.826
3	$v_{info} + c^* + p_{s\_1} + p_{s\_1}$	**0.865**	**0.859**	**0.859**

5 Conclusion

In this work, we introduce a novel framework for the Chinese word semantic relation classification. This framework utilizes the semantic dictionary, linguistic knowledge and word embedding. The results on NLPCC-2017 shared task 1 have demonstrated the efficiency of our approach. We rank No.1 with the result of F1 score 0.859. Our method can be a new clue to other NLP tasks. A promising future work for us is to extend our model to larger training dataset. And another interesting research point is to find a reasonable combination way of features, which maybe bring more excellent performance.

References

1. Girju R, Nakov P, Nastase V, et al.: Classification of Semantic Relations between Nominals [J]. Language Resources & Evaluation, 43(2):105-121 (2009)
2. Hendrickx I, Su N K, Kozareva Z, et al.: SemEval-2010 task 8: multi-way classification of semantic relations between pairs of nominal. The Workshop on Semantic Evaluations: Recent Achievements and Future Directions. Association for Computational Linguistics, 94-99 (2009)
3. Mei, J.J., Zhu, Y.M., et al.: Tongyici cilin. Shanghai: Shanghai Lexicon Publishing Company (1983)
4. Dong, Z. and Dong, Q.: HowNet and the Computation of Meaning, 85-95. Singapore: World Scientific (2006)
5. Bunescu R C, Mooney R J.: A shortest path dependency kernel for relation extraction[C]// Conference on Human Language Technology and Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 724-731 (2005)
6. Zhang M, Zhang J, Su J, et al.: A Composite Kernel to Extract Relations between Entities with both Flat and Structured Features. International Conference on Computational Linguistics and, Meeting of the Association for Computational Linguistics, Proceedings of the Conference, Sydney, Australia, 17-21 July. DBLP (2006)
7. Mikolov, T., Chen, K., Corrado, G. and Dean, J.: Efficient estimation of word representations in vector space. In Proceedings of Workshop at ICLR (2013a)
8. Mikolov, T., Sutskever, I and et al.: Distributed representations of words and phrases and their compositionality. In Proceedings of NIPS, 3111-3119 (2013b)
9. Levy, O. and Goldberg, Y.: Neural word embedding as implicit matrix factorization. In Advances in neural information processing systems, 2177-2185 (2014)

10. Pennington, J. , Socher, R. and Manning, C.D. Glove: Global vectors for word Representation. In Proceedings of EMNLP, 14, 1532-43 (2014)
11. Joseph Turian, Lev Ratinov, Yoshua Bengio.: Word representations: A simple and general method for semi-supervised learning. Proceeding. ACL'10 Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Pages 384-394 (2010)
12. Mrk_si_c, N., S_eaghdha, D. _ O., et al.: Counter-fitting Word Vectors to Linguistic Constraints. arXiv preprint arXiv:1603.00892 (2016)
13. Nguyen, K.A., Walde, S.S.I. and Vu, N.T.: Integrating Distributional Lexical Contrast into Word Embeddings for Antonym-Synonym Distinction. arXiv preprint arXiv:1605.07766 (2016)
14. Chen, Z., Lin, W., et al.: Revisiting word embedding for contrasting meaning. In Proceeding of ACL, 106-115 (2015)
15. Rothe, S. and Sch utze, H.: AutoExtend: Extending Word Embeddings to Embeddings for Synsets and Lexemes. In Proceedings of the ACL-IJNLP, 1793-1803 (2015)
16. Faruqui, M. and Dodge, J., et al.: Retro_tting Word Vectors to Semantic Lexicons. In Proceedings of NAACL (2015)
17. Hashimoto K, Stenetorp P, Miwa M, et al.: Task-Oriented Learning of Word Embeddings for Semantic Relation Classification [J]. Computer Science (2015)
18. T.Mikolov, W.T. Yih, G. Zweig.: Linguistic Regularities in Continuous Space Word Representations. NAACL HLT (2013)
19. Yun fang, Wu. and Minghua Zhang. Overview of the NLPCC 2017 Shared Task: Chinese Word Semantic Relation Classification. In the 6th Conference on Natural Language Processing and Chinese Computing. 2017. Dalian, China.