Relation Classification of Non-saturated Compound Sentences in Chinese via Shallow Convolutional Neural Network

Anonymous ACL submission

Abstract

Most sentences in Chinese writing are compound sentences. Recognition of relation categories is screening for semantic relation of clauses in a compound sentence, and it is the key to analyze the meaning of the whole compound sentences. Non-saturated compound sentences, in which relation words are absent, cannot be identified and classified by the rule of the relation word collocations. In this paper, we identify the relation categories of compound sentences in an unbalanced corpus of non-saturated compound sentences with two clauses. We depart from the traditional approaches which use complicated linguistic knowledge by introducing a convolutional neural network for relation classification that automatically learns features from two clauses and minimizes the dependence on linguist rule. Our model takes advantages of multiple filters with varying window sizes and pre-trained word embeddings, combining the feature of relation words. The experimental results show that our proposed model outperforms the best baseline systems with sentence level features according to different linguistic rule.

1 Introduction

Compound sentences are usually composed of two or more clauses that are semantic closely related, and relation marks (words) are the grammatical component of the compound sentence that is used to connect each clause. In a Chinese compound sentence, however, the relation marks in the clause can be omitted. A Non-saturated compound sentence refers to the compound sentence in which a relation mark does not appear completely. Relation classification of compound sentence refers to the category of semantic relations between two clauses, and compound sentences with multiple clauses can be divided into two clauses. Therefore, we consider compound sentences with two clauses as the object of the study in this paper.

A compound sentence S with the two clauses c1, c2 is defined as a two-tuple $< c1, c2 >$.A non-saturated compound sentences should meet the conditions:

$$(r(c1) = \phi \wedge r(c2) \neq \phi) \vee (r(c1) \neq \phi \wedge r(c2) = \phi)$$

r(c)represents relation mark of the clauses.

The famous Chinese linguist, Prof. Fuyi Xing proposed a three-class system of compound sentences (Xing, 2001). In this paper, the three-class system of compound sentences is used, which includes Expansion, Contingency and Comparison. Each category is divided into some sub-categories. $R < c1, c2 >$ represents Chinese compound sentence relation category.

$$R < c1, c2 > = \{Expansion, Contingency, Comparison\}$$

$$Expansion = \{expansion | coherence | progressive \mid preference\ categories\}$$

$$Contingency = \{contingency \mid recurrence \mid hypothetical \mid conditional \mid target\ categories\}$$

$$Comparison = \{comparison \mid concession \mid denial\ of\ expectation\ categories\}$$

In a compound sentence, there is a strong relationship between the relation classification of compound sentences and the collocation of relation words. A relation mark collocated with different relation marks may conduct different relations of the clauses. For a saturated compound sentence with explicit appearance of a relation word in each clause, the corresponding category can be judged according to the collocation of the relation words. For example, the relation words, “也” (ye, also), can collocate different relation marks such as “既” (ji, both), and “如果” (ru guo, if). In example (S1), the two clauses demonstrate an Expansion relation. (S2): Contingency relation. In non-saturated compound sentences (S3), (S4) and (S5), the relation word “也” (ye, also) appear in the clauses, but the category of compound sentences is difficult to identify.

(S1) 这种电话 既 是用于紧急情况下的报警 (”The phone is used both for emergency alarms”), 也 用于遇到一般困难时的求助 (”And for help in times of general difficulties”)。

(S2) 如果 技术掌握得当 (”If the technology is properly controlled”), 阔叶树移栽 也 有成功的实例 (”Broad-leaved tree transplants also have successful cases”)。

(S3) 当我的独创产品成为世界一流时 (”When my original product became world-class”), 我 也 自然而然跻身于世界强人之列 (”I am also naturally among the world’s best”)。

(S4) 他听课 (”He attended the class”), 也 不打声招呼 (”He also did not say it in advance”)。

(S5) 条件不同 (”The conditions are different”), 面临的任务 也 不同 (”The task is also different”)。

Various approaches have been explored to solve the problem of the relation classification of compound sentences. The most representative method for relation identification is feature-based method, which uses lexical and syntax features that are extracted after performing sentence analysis (Zhou and Yuan, 2008; Huang and Chen, 2012; Li et al., 2014). Yang et al. (2017) introduced semantic relation of clauses, and combined semantic relevance calculation to automatic relation classification.

In this paper, a shallow convolutional neural network (SCNN) model is used to extract related features, grammatical features and semantic features of compound clauses, and identify the category of non-saturated compound sentences in order to reveal the implication of compound sentences.

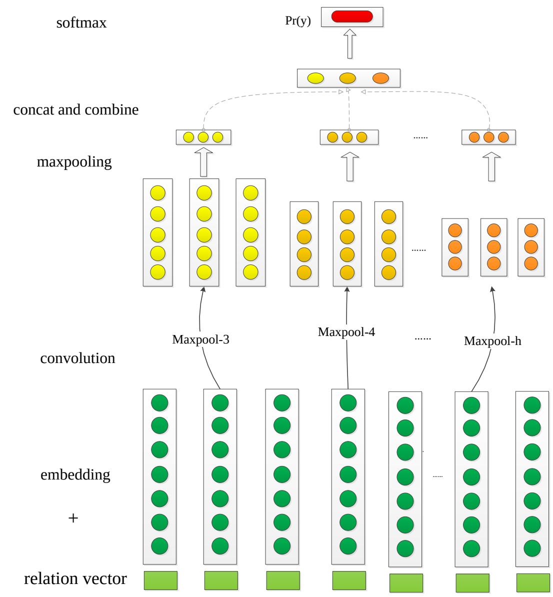

Figure 1: Architecture of shallow convolutional neural network.

2 Methods

The architecture of our SCNN model is illustrated in Figure 1. Firstly, a look-up table is utilized to fetch the embedding of word, forming compound sentences embedding which will be the input of the convolutional layer. In particular, relation word feature is considered in our model. Through the convolution and max-pooling operations, multiple features are obtained by using multiple filters with varying window sizes. Finally, these features will be combined and sent to the final softmax layer after concatenated.

2.1 Relation-Vector and Word-Embedding

A many-to-many relationship exists between relation mark and relation category of compound sentences. On one hand, one relation marker may correspond to different relation categories in different contexts. On the other hand, one relation category may be represented by different relation markers. In

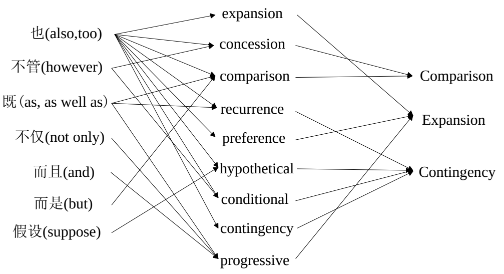

Figure 2: Corresponding examples of word marks and relation classification.

this paper, a relational table with 665 relation words was used. Each relation word in the table may correspond to multiple relation categories. As shown in Figure 2. According to the relation table, each word is discretized by One-Hot Encoding, and the corresponding relation vector $Rv_i \in R^l$ is obtained, where l is the dimension of relation vector. Let $X_i \in R^k$ be the k-dimensional word vector derived from pre-trained word embedddig. Concatenate the two vector into Y_i :

$$Y_i = Rv_i \oplus X_i \tag{1}$$

Where $\oplus$ is the concatenation operator. A sentence of length n (padded where necessary) is represented as

$$Y_{1:\mathrm{n}} = Y_1 \oplus Y_2 \oplus \ldots Y_n \tag{2}$$

In general, let $Y_{\mathrm{i:i+j}}$ refer to the concatenation of vectors $Y_i, Y_{\mathrm{i+1}}, \ldots, Y_{\mathrm{i+j}}$.

2.2 Convolution and Max-pooling

A convolution operation involves a filter $W \in R^{\mathrm{hk}}$, which is applied to a window of h words to produce a new feature (Kim, 2014). For example, a feature c_i is generated from a window of $Y_{\mathrm{i:i+h-1}}$. Filter matrices $[W_1, W_2, \ldots, W_{\mathrm{n-h+1}}]$ utilized to perform the convolution operations for the sentence embedding. The sentence embedding will be transformed to a feature map.

$$c_i = [\ldots, relu(W \cdot Y_{\mathrm{i:i+h-1}} + b), \ldots] \tag{3}$$

Here $[i : \mathrm{i+h-1}]$ indexes the convolutional window and $b \in R$ is a bias term. Additionally, a wide convolution operation between embedding layers and filter matrices was apptied, because it ensures that all weights in the filters reach the entire sentence, including the words at the margins. We then apply a max-over-time pooling operation (Collobert et al., 2011) in the feature map and take the maximum value $\widehat{c} = max(c_i)$ as the feature corresponding to this particular filter. The idea is to extract the most important feature for each feature map.

2.3 Concatenating

Now adding the superscripts and considering different filters $[\widehat{c}^1, \widehat{c}^2, \ldots, \widehat{c}^h]$, they are concatenated to form the sentence representation vector v as below:

$$v = \widehat{c}^1 \oplus \widehat{c}^2 \oplus \cdots \oplus \widehat{c}^h \tag{4}$$

The features are passed to fully concatenated softmax layer whose output is the probability distribution over labels.

2.4 Training

During the period of training, we experimented with two types of word vectors. One is pre-trained word vectors throughout training data, another vector combined pre-trained word vectors and relation word vectors.

The network is trained with mini-batches by back propagation and the gradient-based optimisation is performed using the Adam Optimizer. At the same time, we combine early stopping (Caruana et al., 2001) and dropout (Hinton et al., 2012) to deal with the serious over-fitting problem. Figure 3 shows that as the training steps increase, the loss rate gradually decreases until it tends to be stable in the verification set. At this point, the SCNN model is iterated to the optimal state. During the period of training, the accuracy rate is improved gradually with the increase of iteration steps. Figure 4 shows that the closer the accuracy is to 1, the stronger the fitting capacity of the model is.

The cross-entropy function is proven to be able to accelerate the back propagation algorithm and provide good overall network performance with relatively short stagnation period (Turian et al., 2010), especially for classification tasks. To construct the objective function, the cross-entropy loss function and an L2 regularization term are considered. We use ReLu as the activation function for our model,

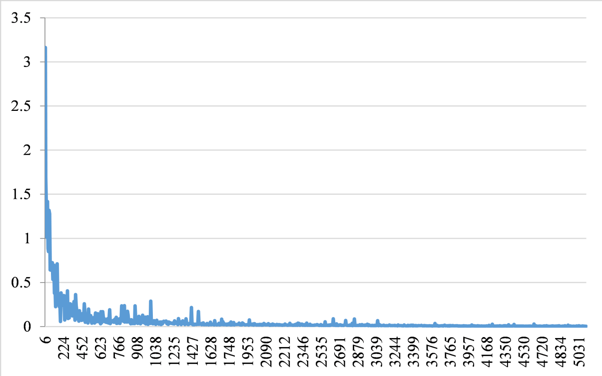

Figure 3: The loss rate decrease as training steps increase

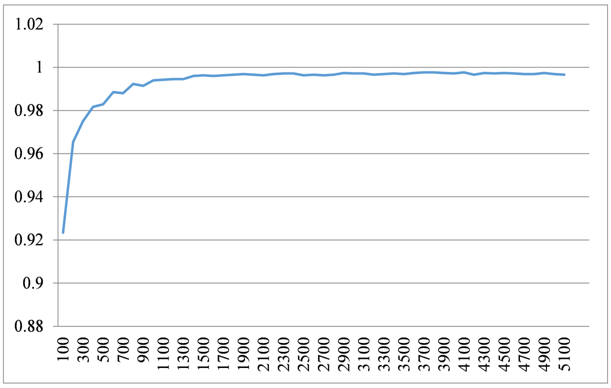

Figure 4: The accuracy rate increases with training steps.

which yields comparable or better results than sigmoid-like functions, but converges faster.

'The Corpus of Chinese Compound Sentence (CCCS[1])' is a corpus which contains 658,447 compound sentences with multiple clauses. Relation marks and sentence relation category are labeled. A Total of 24,706 sentences are elaborately selected from the CCCS, which form a corpus of Non-saturated Chinese Compound Sentence with two clauses. This corpus is abbreviated as NCCST.

3 Experiments and Results

NCCST was divided into training set and test set with the ratio of 7:3, the distribution of compound sentences relations is shown in Table 1. The imbalanced corpus will seriously affect the training effect of the model, it can easily induce overfitting and lead to converaging difficulty of the model. Over-sampling techniques (He and Garcia, 2009) that duplicating instances from the minority classes until a balanced distribution was used on the imbalanced data increased the SCNN performances to that of the SCNN trained with balanced data. We duplicate the samples of Comparison, Contingency and Expansion in the training sets proportionally to derive balanced training data. Ultimately, the number of all categories reached 11,652.

We conducted three experiments. The first was to input pre-trained word embedding into SCNN model. The second was to add hand-crafted word marks feature into pre-trained word embedding as input of SCNN model, called FCNN. The third experiment was to compare the Decision Tree Classifier for relation classification of Chinese sentence proposed by Huang and Chen (2012). It should be noted, however, that we also conducted an evaluation on the stop word[2], and found that its effect is small. The experimental results are shown in Table 2.

Table 2 summarizes the performances of models SCNN, FCNN and C5.0 (Huang and Chen, 2012) in precision, recall, accuracy and MicroF1. For fair comparisons, all the competitor models were trained on the same training set. We can see that the both of our shallow CNN and FCNN outperforms C5.0 by at least 16.84%, indicating that our treatment is better than previous conventions in capturing syntactic structures for relation classification. It was noted that FCNN, with extract considerations for relation marks, outperforms SCNN. We found that the FCNN improve 0.32% in MacroF1 scores and 0.27% in accuracy, showing that relation marks offer more discrimination information that benefits the relation classification task.

4 Discussions

A statistical analysis of the results shows that the FCNN model has a good recognition effect on the sentences in which relation word only conducts one relationship category. Moreover, for those sentences one relation word such as '也 (ye, also)', can lead various relations (as shown in figure 2), FCNN is also more accurate than the SCNN. For the examples mentioned, FCNN could recognize S1 as Expansion rela-

[1]http://218.199.196.96:8080/jiansuo/TestFuju.jsp

[2]top word that do not include relation words

Relations	Training set	Test set	Total	Percentage
Comparison	4857	2083	6940	28.09%
Contingency	11657	4995	16652	67.40%
Expansion	781	333	1114	4.51%

Table 1: Statistic of the data set.

Methods	Precision	Recall	Accuracy	MacroF1
C5.0	88.80%	72.80%	92.20%	80.00%
SCNN	90.74%	88.71%	96.88%	96.84%
FCNN	90.56%	91.07%	97.15%	97.16%

Table 2: the experimental results of proposed method against other models.

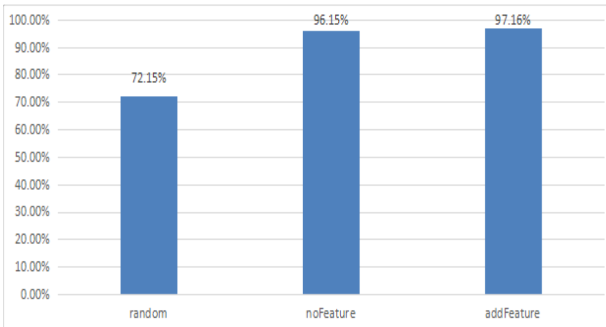

Figure 5: Effects of different inputs.

tion, S2 as Comparison relation and S3 as Contingency relation. The results proved that the feature of the relation word plays a key role in the relation recognition of compound sentences.

We further analyzed the influence of using pre-trained word embedding for initialization. Figure 5 shows that using pre-trained word embedding can obtain around 24%, compared with the result using random initialization with a uniform distribution in the range of [-1,1].

Our neural model performs as well as or even outperforms the systems with lexical and syntax features across different linguistic rule. Convolutional neural network models are an attractive alternative for this task, and neural network can automatically learn some semantic features without useing existing parsing tools. On one hand, the method based on feature strongly depends on the quality of feature extracting, and it is time-consuming. On the other hand, extracting features is often derived from the output of pre-existing natural language processing (NLP) systems, which leads to the propagation of the errors in the existing tools and hinders the performance of these systems. Moreover, our model worked efficiently during training.

5 Conclusions

In this paper, we learned more effective relation representations for relation classification with a convolutional neural network model. We further presented a FCNN classifiers approach to handle the semantic relation in two clauses of Compound sentences better. The model is effective for compound sentence relation recognition. The system is improved when relation mark features are added. The automatically learned features yield excellent results and can replace the elaborately designed features that are based on prior linguistics knowledge and outputs of existing NLP tools.

In the future, we will build neural network models on dependency syntax trees to extract more grammar features. Besides, relation classification of Chinese compound sentences with more clauses and no explicit relation marks will be further explored.

References

Rich Caruana, Steve Lawrence, and C Lee Giles. 2001. Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping. In *Advances in neural information processing systems*. pages 402–408.

Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. 2011. Natural language processing (al-most) from scratch. *Journal of Machine Learning Research* 12(Aug):2493–2537.

Haibo He and Edwardo A. Garcia. 2009. Learning from imbalanced data. *IEEE Trans. Knowl. Data Eng.* 21(9):1263–1284. https://doi.org/10.1109/TKDE.2008.239.

Geoffrey E. Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2012. Improving neural networks by preventing co-adaptation of feature detectors. *CoRR* abs/1207.0580. http://arxiv.org/abs/1207.0580.

Hen-Hsen Huang and Hsin-Hsi Chen. 2012. Contingency and comparison relation labeling and structure prediction in chinese sentences. In *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Association for Computational Linguistics, pages 261–269. http://www.aclweb.org/anthology/W12-1636.

Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, pages 1746–1751. https://doi.org/10.3115/v1/D14-1181.

Yancui Li, Jing Sun, Guodong Zhou, and Wenhe Feng. 2014. Recognition and classification of relation words in the compound sentences based on tsinghua chinese treebank. *Acta Scientiarum Naturalium Universitatis Pekinensis* 50(1):118–124.

Joseph P. Turian, Lev-Arie Ratinov, and Yoshua Bengio. 2010. Word representations: A simple and general method for semi-supervised learning. In *ACL 2010, Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, July 11-16, 2010, Uppsala, Sweden*. pages 384–394. http://www.aclweb.org/anthology/P10-1040.

Fuyi Xing. 2001. *The research of Chinese compound sentence*. The Commercial Press, Beijing, China.

Jincai Yang, Zhongzhong Chen, Xianjun Shen, and Jinzhu Hu. 2017. Automatic recognition of relation category of non-saturated compound sentences with two clauses. *Application Research of Computers* 34(10):2950–2953.

Wencui Zhou and Chunfeng Yuan. 2008. Automatic recognizing of complex sentences with coordinating relation. *Application Research of Computers* 25(3):764–767.