Attention-Based Convolutional Neural Networks for Chinese Relation Extraction

Wenya Wu, Yufeng Chen*, Jinan Xu, Yujie Zhang

School of Computer and Information Technology, Beijing Jiaotong University
{wuwy, chenyf, jaxu, yjzhang}@bjtu.edu.cn

Abstract. Relation extraction is an important part of many information extraction systems that mines structured facts from texts. Recently, deep learning has achieved good results in relation extraction. Attention mechanism is also gradually applied to networks, which improves the performance of the task. However, the current attention mechanism is mainly applied to the basic features on the lexical level rather than the higher overall features. In order to obtain more information of high-level features for relation predicting, we proposed attention-based piecewise convolutional neural networks (PCNN_ATT), which add an attention layer after the piecewise max pooling layer in order to get significant information of sentence global features. Furthermore, we put forward a data extension method by utilizing an external dictionary HIT IR-Lab Tongyici Cilin (Extended). Experiments results on ACE-2005 and COAE-2016 Chinese datasets both demonstrate that our approach outperforms most of the existing methods.

Keywords: Relation Extraction, Convolutional Neural Networks, Attention Mechanism.

1 Introduction

Relation extraction is one of the basic tasks of Natural Language Processing (NLP), and it is used to identifying the semantic relation holding between two nominal entities in a sentence. It converts unstructured data into structured data. Moreover, relation extraction provides significant technical support for many NLP tasks, such as mass information processing, Chinese information retrieval, knowledge base automatic construction, Machine Translation and automatic abstracting. At present, the scale of Chinese information has increased rapidly, and the technical requirements for Chinese texts processing have been gradually improved. However, there are relatively scarce researches about neural networks in Chinese relation extraction. Therefore, this paper focuses on Chinese corpus.

In the past ten years, deep neural networks have achieved the best results in various fields. For example, models CNNs-based (Zeng et al., 2014; Xu et al., 2015; Dos Santos et al., 2015) and RNNs-based (Hashimoto K. et al., 2013; Zhou et al., 2016) have got fairly competitive results in relation extraction task. Methods that combine CNNs with RNNs (Liu et al., 2015; Cai et al., 2016) have emerged. Attention mechanism (Zhou et al., 2016; Wang et al., 2016) is also frequently used in network structure. Experiments

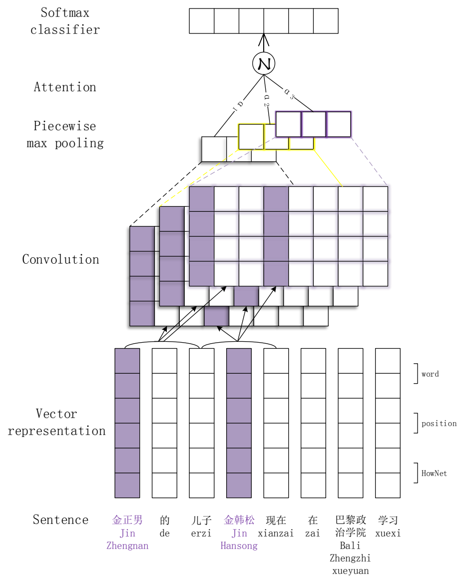

Fig. 1. The architecture of PCNN_ATT used for relation extraction, illustrating the procedure for handling one instance and predicting the relation between "金正男(Jin Zhengnan)" and "金韩松(Jin Hansong)".

show that attention mechanism often can improve performance. Regrettably, most attention mechanisms focus on the low-level features like lexical characteristics. Although they are well-interpreted and achieve better results, we should also pay attention to more high-level features, such as the overall advanced features obtained after convolution and max pooling.

Therefore, this paper proposes novel neural networks PCNN_ATT (Attention-based piecewise convolutional networks) for relation extraction. In this model, we choose a convolutional neural network (CNN) to extract features, since CNN is proficient in modeling flat structure and can generate a fixed-size vector with the most meaningful features (Liu et al., 2015). As illustrated in Fig. 1, we employ a convolution layer to extract the semantic features of sentences. Afterwards, to obtain more abundant information, we adopt a piecewise max pooling method presented by Zeng et al. (2015) to further extract high-level features. In the convolutional layer, we use multiple convolution kernels to obtain different types of text features. We believe that different features obtained by different kernels contribute differently to the final relation. Therefore, We add an attention layer to allocate weights effectively. Finally, we extract relation between two nominal entities with the relation vector produced by a softmax classifier.

The main contribution of this model is using PCNN with an attention mechanism, which could automatically focus on the high-level features obtained by max pooling that have decisive effects on relation predicting. In addition, we add a HowNet embedding of the entities in vector representation layer by using HowNet resource to express lexical features better. We evaluate the model on COAE2016 and ACE2005 datasets in relation extraction task. The experimental results show that this model is competitive.

Otherwise, recently most models about deep learning have achieved good results of relation extraction for English, while the effects on Chinese data sets are not so good because of the lack of Chinese annotation data (Sun et al., 2017). To tackle this problem, we use the HIT IR-Lab Tongyici Cilin to expand the existing corpus. The specific method and experiments will be introduced in Section 4.3.

The major contributions of this paper are as follows.

- A novel CNN architecture is proposed, which relies on a fancy attention mechanism to capture high-level features (obtained after pooling the results of different convolution kernels) attention. The experiments on COAE2016 and ACE2005 datasets show that our PCNN_ATT model on relation extraction task outperforms other state-of-the- art models.
- The shortest dependency path is utilized to increase the scale of Chinese dataset COAE2016, and we expand the corpus efficiently by taking advantage of external semantic dictionary Tongyici Cilin, which makes it more suitable for deep learning method.
- Hypernym features of nominals in the HowNet are integrated into our model. The experimental results on ACE2005 datasets demonstrate the validity of those features.

2 Related Work

It's well known that relation extraction is one of the important tasks of information extraction. In recent years, many scholars and experts in the Natural Language Processing field have devoted themselves to the construction of Knowledge Graph that makes search more depth and breadth. Information extraction is a key step in the construction processes of Knowledge Graph.

The methods of relation extraction include supervised method, unsupervised method, semi-supervised method and open domain relation extraction method. Compared with other methods, supervised method has more abundant researches. Generally speaking, relation extraction typically cast as a standard supervised multi-class or multi-label classification task. Depending on the representation of relation instances, these supervised means can be further divided into feature-based and kernel-based.

Feature-based methods usually need artificial design characteristics that typically express corpus property. For example, Huang et al. (2010) combined the basic features of words, entities and grammar. Unfortunately, it is not easy to design effective features artificially, which will cost lots of time and effort. Different from feature-based approach, kernel-based method (Bunescu and Mooney 2005) does not require the construction of feature vectors. It directly uses the original form of the string as the processing object, and then computes the similarity function between any two objects. However, the fatal disadvantage of this method is that the speeds of training and prediction are too slow.

Over the past decade, deep learning has shown its outstanding characteristics. Socher et al. (2012) proposed the model named Recursive Matrix-Vector Model (MV-RNN) that learns semantic compositionality from syntactic trees. To capture important information appeared at any position in the sentence, Zhou et al. (2016) put forward Attention-Based Bidirectional Long Short-Term Memory Networks (Att-BLSTM). Att-BLSTM can not only consider the past and future information of the sentence, but also determine the most pivotal information. In addition, CNN model also be used for relation extraction. Convolutional Deep Neural Network (CDNN) was used for relation classification by Zeng et al. (2014) , and CDNN adopted position feature and WordNet hypernyms of nouns feature. In our work, position features also be utilized and we employ HowNet hypernyms of entities to make better performance. Zeng et al. (2015) proposed piecewise convolutional neural networks (PCNN) to alleviate the influence of noise caused by the feature extraction process, which are also adopted in this paper. Wang et al. (2016) employed a CNN architecture relying on a novel multi-level attention mechanism to capture both entity-specific attention and relation-specific pooling attention. Inspired by Wang et al. (2016), our work also adds an attention layer to capture more important information of high-level features.

3 Methodology

In this section, we will introduce the neural network architecture used in our paper. Figure 1 illustrates the procedure that handles one instance. This procedure includes five main parts: **Vector representation**, **Convolution**, **Piecewise max pooling**, **Attention** and **Softmax output**. We describe these parts below.

3.1 Vector Representation

The input embedding of the model consists of three parts: word embedding, position embedding and HowNet hypernym embedding.

Word Embedding. Word embeddings are distributed representations of words that map each word in texts to a low-dimensional vector. Given a sentence consisting of T words $S = [x_1, x_2, \cdots, x_T]$, every word x_i is converted into a real-valued vector e_i by looking up word embeddings. So $WE = [e_1, e_2, \cdots, e_T]$.

Position Embedding. To specify the target nouns in the sentence, Zeng et al.(2014) employed Position Feature(PF) that is the combination of the relative distances of the current word to ent_1 and ent_2.(ent_1 and ent_2 represents entity1 and entity2, respectively.) Figure 2 shows an example of the relative distance. The relative distance from word *儿子(erzi)* to *金正男(Jin Zhengnan)* and *金韩松(Jin Hansong)* are 2 and -1. So the position embedding of the sentence S is $PE = [\, pe_1, pe_2, \cdots, pe_T \,]$, and $pe_i(i = 1,2, \cdots T) = [d_{i1}, d_{i2}]$.($d_{i1}$ and d_{i2} represent the distance vectors for the i^{th} word of S to ent_1 and ent_2.)

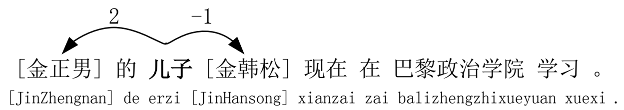

Fig. 2. An example of relative distances.

HowNet Hypernym Embedding. HowNet is a network of semantic relationships among Chinese words. In HowNet, the formal description of words is organized in three layers: “word”, “concept”, and “sememe”. Every word can be described by several concepts. And “concept” is described by a kind of knowledge expressing language, which is composed by sememes. “Sememe” is the basic semantic unit, all the sememes are organized into a hierarchical tree structure by the Hypernym-Hyponym relations. We use HowNet 2008 in our experiments, there are 1,700 sememes, 28,925 concepts, and 96,744 Chinese words.

The hypernym has been used to enhance the feature quality. For instance, WordNet usually be employed to perform better results(Rink and Harabagiu, 2010; Zeng et al., 2014). HowNet contains multiple relations between Chinese words, including hypernyms. So for the entities in a sentence, we add its HowNet hypernym as a feature. The HowNet Hypernym Embeddings of S is HHE = $[\mathrm{hhe}_1, \mathrm{hhe}_2, \cdots, \mathrm{hhe}_T]$.

$$\mathrm{hhe}_i = \begin{cases} e_{x_i\mathrm{HowNetH}}, & \text{if } x_i \text{ is entity,} \\ 0, & \text{if not.} \end{cases} \tag{1}$$

Where $e_{x_i\mathrm{HowNetH}}$ is the word embedding of entity x_i hypernym in HowNet. **0** is a vector that has the same dimension as $e_{x_i\mathrm{HowNetH}}$.

We concatenate the word representation, position representation and HowNet Hypernym representation as the input Emb_S of the network, and $\mathrm{Emb}_S = [\mathrm{WE}, \mathrm{PE}, \mathrm{HHE}]^T$.

3.2 Convolution

Collobert et al.(2011) considered convolution approach can merge all local features to perform prediction globally. Convolution is an operation between a vector of weights w and a vector of inputs that is treated as a sequence. The weights matrix w is regarded as the filter for the convolution. The ability to capture different features typically requires the use of multiple filters in the convolution. Under the assumption that we use n filters ($W = \{w_1, w_2, \cdots, w_n\}$). The convolution result is a matrix $C = \{c_1, c_2, \cdots, c_n\}$.

3.3 Piecewise Max Pooling

To capture the structural information between two entities better, we employ piecewise max pooling proposed by Zeng et al.(2015) instead of single max pooling. As shown in Figure 1, the output of each convolutional filter c_i is divided into three segments

$\{c_{i1}, c_{i2}, c_{i3}\}$ by 金正男(Jin Zhengnan) and 金韩松(Jin Hansong). The piecewise max pooling procedure can be expressed as follows:

$$p_{ij} = \max(c_{ij}) \qquad 1 \le i \le n,\ \ 1 \le j \le 3 \tag{2}$$

So ,$p_i = [p_{i1}, p_{i2}, p_{i3}]$. We concatenate all vectors that obtained by piecewise max pooling on the outputs of convolutional filters as $P = [p_1, p_2, \cdots, p_n]$.

3.4 Attention

Attentive neural networks have recently achieved good results in relation extraction. For example, Zhou et al. (2016) added an attention layer after Bidirectional LSTM layer. And Wang et al. (2016) proposed a novel multi-level attention mechanism to capture both entity-specific attention and relation-specific pooling attention.

What is different from them is that our attention mechanism automatically focus on high-level global features obtained by convolution and pooling multiple filters that have decisive influence on prediction relations. Every convolution filter will extract a kind of high-level feature about sentence globally, and different features have different contributions to predict entity relation. Therefore, we add an attention layer to find the most advantageous features for relation predicting. The representation r of the sentence S is produced by the following formula:

$$M = \tanh(P) \tag{3}$$

$$\alpha = \mathrm{softmax}(w^T M) \tag{4}$$

$$r = P\alpha^T \tag{5}$$

where $P \in R^{3\times n}$, n is the number of filters, w is a trained parameter vector and w^T is a transpose.

We obtain the final sentence representation used for predicting relation from:

$$h^* = \tanh(r) \tag{6}$$

3.5 Softmax Output

In this part, we employ a softmax classifier to predict label $\hat{y}$ from a discrete set of classes Y for a sentence S. And h^* is involved in the following formula as input:

$$\hat{p}(y|S) = \mathrm{softmax}(Wh^* + b) \tag{7}$$

$$\hat{y} = \arg\max_y \hat{p}(y|S) \tag{8}$$

We combine a dropout with L2 regularization to alleviate overfitting in our paper.

In this section, we have introduced each part of the model in detail, especially for our innovations, attention layer and HowNet embedding in vector representation layer.

4 Experiments

4.1 Datasets and Evaluation Metrics

COAE2016. The 8th Chinese Orientation Analysis and Evaluation(COAE2016) increased the task of relation classification for knowledge extraction. Our experiments are conducted on the dataset provided in this task, and the dataset consists of a training set of 988 sentences and a test set of 483 sentences. In this task, there are ten types of relation, as shown in table 1.

ACE2005. ACE2005 Chinese corpus is divided into three categories: Newswire (NW), Broadcast News (BN), and Weblog (WL). Table 2 shows the distribution of the data. The corpus consists of data of various types annotated for entities and relations. The relation types are divided into 6 categories and 18 subcategories. Table 3 shows the distribution of the number for 6 categories. We randomly select 2/3 as training set and 1/3 as test set.

We use the macro-averaged Precision, Recall and F1-score to evaluate our systems.

Table 1. Relation type table.

Relation name	Relation symbol
Date of birth of PER	Cr2
Birthplace of PER	Cr4
Graduate institution of PER	Cr16
Spouse of PER	Cr20
Children of PER	Cr21
Senior executive of ORG	Cr28
Employees' number of ORG	Cr29
Founder of ORG	Cr34
Founding time of ORG	Cr35
Headquarters of ORG	Cr37

Table 2. Data source distribution table.

	Chars	Files	Rate(%)
NW	121797	238	40
BN	120513	298	40
WL	65681	97	20
Total	307991	633	100

Table 3. Data relation type distribution table.

Category	ART	PART-WHOLE	PHYS	ORG-AFF	PER-SOC	GEN-AFF	Total
number	507	1862	1360	1874	486	1660	7749

4.2 Parameter Settings

Collobert et al.(2011) proved that word embedding learned from a large number of unlabeled data are far more satisfactory than the randomly initialized embeddings. Therefore, we use word embeddings pre-trained in 13 million words People's Daily corpus in our experiments. We tune all of the models using three-fold validation on the training set. We select the dimension of word embedding d_w among {50, 100, 200}, the dimension of position embedding d_p among {5, 10, 20}, the dimension of HowNet hypernym embedding d_h among {5, 10, 20}, the windows size w among {3, 5, 7}, the number of filters n among {100, 150, 200, 230, 250}, the learning rate λ among {0.001, 0.01, 0.1, 1.1}, while the batch size k among {50, 100, 150, 200}. The best configurations are: $d_w = 100, d_p = 5, d_h = 10, w = 3, n = 230, \lambda = 0.01, and\ k = 50.$ According to experience, the dropout rate is fixed to 0.5.

4.3 Corpus Expansion for COAE2016

As is well-know that large-scale labeled data is essential for deep learning methods. However, manually tagging large amounts of data is time consumed and vigour cost, and the existing Chinese labeled corpus is relatively scarce. Therefore, the effective way of data expansion brings new gospel. In this section, we will introduce how we use HIT IR-Lab Tongyici Cilin (Extended) and SDP to expand limited Chinese labeled corpus like COAE2016.

HIT IR-Lab Tongyici Cilin(Extended). We use the HIT IR-Lba Tongyic Cilin (Hereinafter referred to as Cilin), which was expanded by the Information Retrieval Laboratory of Harbin Institute of Technology. Cilin has increased the number of entries from 39,099 to 77,343 and it was organized according to the tree hierarchical structure.

Cilin encodes the words by category. Table 4 shows the method of word encoding. The fifth-level classification results can be divided into three specific situations, e.g., some lines are synonyms, some lines are related words, and some lines have only one word. That is to say, there are three markers in eighth places, namely "=", "#" and "@". "=" represents equality and synonym. "#" represents unequal and similar. "@" indicates self-sealing and independence, and it has neither synonyms nor related words in the dictionary. For example, Cb02A01= 东南西北 四方. "东南西北(dongnanxibei)"and "四方(sifang)" have the same meaning. In our method, we mainly use synonyms.

Shortest Dependency Path. The dependency grammar reveals sentence syntactic structure by parsing the dependencies among the components in the language units. Intuitively speaking, grammatical components such as subject-predicate-object and attribute-adverbial-complement are identified in dependency syntactic analysis, and the relationships among the components are analyzed. As shown in Figure 3, we perform dependency parsing on the sentence.

Table 4. Cilin words coding table.

Coding position	1	2	3	4	5	6	7	8
Symbol example	D	a	1	5	B	0	2	=\#\@
Symbolic nature	Big class	Middle class	Small class		Word group	Atomic word group		
Level	Level 1	Level 2	Level 3		Level 4	Level 5		

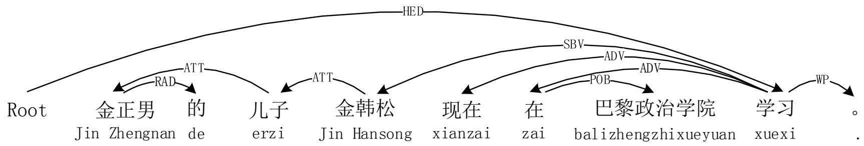

Fig. 3. Dependency parsing result

In succession, we find the shortest path between two entities in the result of dependency parsing. According to the algorithm, the shortest path between “金正男(Jin Zhengnan)” and “金韩松(Jin Hansong)” is as follows:

金正男 Jin Zhengnan ←ATT— 儿子 erzi ←ATT— 金韩松 Jin Hansong

Corpus Expansion. In this part, we will describe the data expansion method in details with give examples. First, we preprocess the data, such as word segmentation. Then, we perform dependency parsing on sentences[1] and find the shortest path of two entities. For words that appear in the shortest path (exclude entities), we find their synonyms in Cilin and replace them in turns to get new data. Table 5 shows some examples.

Since Chinese word polysemy often occurs, we calculate the similarity between the word replaced and its synonyms. If the similarity exceeds the threshold, we replace it. We use Similarity Model proposed by Li et al. (2017) to compute word similarity.

4.4 Experimental Results and Analysis

Table 6 shows the experimental results on the COAE2016 dataset. We compare with previous approaches. In the case of a small data set (988 sentences for training), SVM method is superior to CNN. Our model PCNN_ATT is obviously better than baselines. It is not only outperforming an SVM-based approach, but also superior the CNN model with a relative improvement of 11.80%, which demonstrate PCNN_ATT have advantages in relation classification task. Our model is superior to PCNN (Zeng et al., 2015) with a relative improvement of 2.62%. Entity types, such as *person*, *location* and *organization*, are helpful to distinguish some relations. In this dataset, relation types are extremely specific. The addition of entity types can largely improve the classification results and solve some misclassification problems. After adding entity type (ET) information, the result is increased by 8.29%. Finally, the experiment was performed on the

[1] We use Stanford Parser to perform dependency parsing on sentences.

Table 5. New data generation sample table

Original sentence	SDP	New data
金正男的儿子金韩松现在在巴黎政治学院学习。(Jin Zhengnan de erzi Jin Hansong xianzai zai bali zhengzhi xueyuan xuexi .)	金正男 <- 儿子 <- 金韩松	1.金正男 的 子嗣(zisi) 金韩松 现在 在 巴黎政治学院 学习 。 2.金正男 的 幼子(youzi) 金韩松 现在 在 巴黎政治学院 学习 。 3.金正男 的 犬子(quanzi) 金韩松 现在 在 巴黎政治学院 学习 。
松田耕平代替父亲松田恒次的位置，成为马自达的会长 。(Songtiangengping daiti fuqin songtianhengci de weizhi , chengwei mazida de huizhang .)	松田耕平 <- 代替 -> 成为 -> 会长 -> 马自达	1.松田耕平 顶替(dingti) 父亲 松田恒次 的 位置 ，成为 马自达 的 会长 。 2.松田耕平 取代(qudai) 父亲 松田恒次 的 位置 ，成为 马自达 的 会长 。 3.松田耕平 代替 父亲 松田恒次 的 位置 ，成了(chengle) 马自达 的 会长 。 4.松田耕平 代替 父亲 松田恒次 的 位置 ，变成(biancheng) 马自达 的 会长 。

Table 6. Comparison of different models on COAE2016 dataset. WV,PF ET stand for word vectors, position features and entity types; *(Ext) stands for results on the extended dataset.

Methods	Features	Macro		
		P	R	F1
SVM	POS，Entity order，Entity distance，Entity context	76.78	64.70	66.29
SVM(Ext)		77.23	65.72	67.96
CNN	WV,PF	60.83	55.60	56.69
CNN(Ext)	WV,PF	62.23	57.75	58.61
PCNN	WV,PF	69.95	61.88	65.67
PCNN(Ext)	WV,PF	71.63	62.31	66.65
PCNN_ATT	WV,PF	75.38	66.89	68.29
PCNN_ATT	WV,PF,ET	77.64	76.23	76.58
PCNN_ATT(Ext)	WV,PF,ET	**79.26**	**78.47**	**78.41**

expanded dataset (11,328 sentences) with the result of 78.41%, increased 1.83%. The results of each model on the extended dataset have been improved, which prove that the data expanded method is effective.

Table 7 shows the experimental results on the ACE2005 dataset. We don't apply corpus extension method on ACE2005, because the scale of this dataset is adequate for model training. As we see that SVM method is much better than CNN method in this task. When we have the same features, our model PCNN_ATT performances well with a relative improvement of 13.95% and it is better than PCNN (Zeng et al., 2015) with a relative improvement of 5.63%. After adding entity type (ET) and entity subtype (ES), the result is raised by 0.5%. Finally, with the addition of HowNet hypernyms, the result is nearly 0.6% higher.

Table 7. Comparison of different models on ACE2005 dataset. WV,PF,ET,ES,HowNet stand for word vectors, position features, entity types, entity subtypes and HowNet hypernyms.

Methods	Features	Macro		
		P	R	F1
SVM	POS, Entity order, Entity type and subtype, Entity context	76.13	70.18	73.27
CNN	WV,PF	59.65	58.26	58.95
PCNN	WV,PF	68.25	66.31	67.27
PCNN_ATT	WV,PF	73.11	73.01	72.90
PCNN_ATT	WV,PF,ET,ES	**73.96**	73.28	73.37
PCNN_ATT	WV,PF,ET,ES,HowNet	73.77	**74.12**	**73.94**

Based on the above experiments, our model PCNN_ATT shows competitive results. Adding an attention mechanism after the piecewise max pooling layer can improve performance. The attention mechanism can assign weights to the high-level features obtained by each convolution kernel, and gives greater weights to the features that contribute to the prediction of relation.

5 Conclusion

In this paper, we propose a novel neural network PCNN_ATT, to improve the performance of relation extraction. The PCNN_ATT model added an attention layer after piecewise max pooling layer, which pays more attention to high-level global features. And we put forward an effective corpus expansion method by utilizing external dictionary HIT IR-Lab Tongyici Cilin to make up insufficient data problem. We demonstrate the effectiveness of our method by evaluating the model on COAE-2016 and ACE-2005 datasets. PCNN_ATT achieves a better performance at capturing more important features, compared with some common neural networks. A significant improvement is observed when PCNN_ATT is used, outperforming most of existing methods.

Acknowledgements. The authors are supported by National Nature Science Foundation of China (Contract 61370130 and 61473294), and the Fundamental Research Funds for the Central Universities(2015JBM033), and International Science and Technology Cooperation Program of China under grant No. 2014DFA11350.

References

1. Zeng, D., Liu, K., Lai, S., Zhou, G., Zhao, J.: Relation classification via convolutional deep neural network. In: Proceedings of COLING, pp. 2335–2344 (2014).
2. Zeng, D., Liu, K., Chen, Y., Zhao, J.: Distant supervision for relation extraction via piecewise convolutional neural networks. In: Proceedings of EMNLP, pp. 17–21. Association for Computational Linguistics, Stroudsburg (2015).

3. Lin, Y., Shen, S., Liu, Z., Luan, H., Sun, M.: Neural relation extraction with selective attention over instances. In: Proceedings of ACL, pp. 2124–2133. Association for Computational Linguistics, Berlin (2016).
4. Jiang, X., Wang, Q., Li, P., Wang, B.: Relation extraction with multi-instance multi-label convolutional neural networks. In: Proceedings of COLING, pp. 1471–1480 (2016).
5. dos Santos, C.N., Xiang, B., Zhou, B.: Classifying relations by ranking with convolutional neural networks. In: Proceedings of ACL (2015).
6. Liu, Y., Wei, F., Li, S., Ji, H., Zhou, M., Wang, H.: A dependency-based neural network for relation classification. Computer Science, (2015).
7. Bunescu, R.C., Mooney, R. J.: A shortest path dependency kernel for relation extraction. In: Proceedings of HLT/EMNLP, pp. 724-731. Association for Computational Linguistics, Vancouver (2005).
8. Wang, L., Cao, Z., Melo, G.D., Liu, Z.: Relation classification via multi-level attention CNNs. In: Proceedings of ACL, pp. 1298–1307. Association for Computational Linguistics, Berlin (2016).
9. Li, S., Xu, J.A., Zhang, Y., Chen, Y.: A method of unknown words processing for neural machine translation using HowNet. In: Proceedings of CWMT, pp. 22-29. Springer, Singapore (2017).
10. Sun, J., Gu, X., Li, Y., Xu, W.: Chinese entity relation extraction algorithms based on COAE2016 datasets. Journal of Shandong University(Natural Science) 52(9), 7-12 (2017).
11. Liu, D., Peng, C., Qian, L., Zhou, G.: The effect of Tongyici Cilin in Chinese entity relation extraction. Journal of Chinese Information Processing 28(2), 91-99 (2014).
12. Cai, R., Zhang, X., Wang, H.: Bidirectional recurrent convolutional neural network for relation classification. In: Proceedings of ACL, pp. 756-765. Association for Computational Linguistics, Berlin (2016).
13. Liu, Y., Wei, F., Li, S., J, H., Zhou, M., Wang, H.: A dependency-based neural network for relation classification. Computer Science (2015).
14. Xu, K., Feng, Y., Huang, S., Zhao, D.: Semantic relation classification via convolutional neural networks with simple negative sampling. Computer Science 71(7), 941-9 (2015).
15. Zhou, P., Shi, W., Tian, J., Qi, Z., Li, B., Hao, H., Xu, B.: Attention-based bidirectional long short-term memory networks for relation classification. In: Proceedings of ACL, pp. 207–212. Association for Computational Linguistics, Berlin (2016).
16. Hashimoto, K., Miwa, M., Tsuruoka, Y., Chikayama, T.: Simple customization of recursive neural networks for semantic relation classification. In: Proceedings of EMNLP, pp. 1372-1376. Association for Computational Linguistics, Seattle (2013).
17. Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., Kuksa, P.: Natural language processing (almost) from scratch. JMLR, 12:2493-2537 (2011).
18. Socher, R., Huval, B., Manning, C.D., Ng, A.Y.: Semantic compositionality through recursive matrix-vector spaces. In: Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 1201-1211 (2012).
19. Rink, B., Harabagiu, S.: UTD: Classifying semantic relations by combining lexical and semantic resources. In: Proceedings of the 5th International Workshop on Semantic Evaluation, pp. 256-259. Association for Computational Linguistics (2010).