

IRCMS at SemEval-2018 Task 7 : Evaluating a basic CNN Method and Traditional Pipeline Method for Relation Classification

Zhongbo Yin, Zhunchen Luo, Wei Luo*, Bin Mao, Changhai Tian, Yuming Ye, Shuai Wu
Information Research Center of Military Science, PLA Academy of Military Science, China
zhongboyin@foxmail.com; {zhunchenluo, lwowen79}@gmail.com;
{miscy210, jamestch, yuming-ye}@163.com; wus1986@gmail.com

Abstract

This paper presents our participation for sub-task1 (1.1 and 1.2) in SemEval 2018 task 7: Semantic Relation Extraction and Classification in Scientific Papers (Gábor et al., 2018). We experimented on this task with two methods: CNN method and traditional pipeline method. We use the context between two entities (included) as input information for both methods, which extremely reduce the noise effect. For the CNN method, we construct a simple convolution neural network to automatically learn features from raw texts without any manual processing. Moreover, we use the softmax function to classify the entity pair into a specific relation category. For the traditional pipeline method, we use the Hackabout method as a representation which is described in section3.5. The CNN method's result is much better than traditional pipeline method (49.1% vs. 42.3% and 71.1% vs. 54.6%).

1 Introduction

Scientific paper, as a major source of new technology, is a common way for tracing the dynamics of a research domain. With lots of papers published every year, scholars can't read all of them to extract useful aspects for our research domain. Information extraction (IE) is a main NLP aspects for analyzing scientific papers, which includes named entity recognition (NER) and relation extraction (RE). Scientific papers' information extraction is identify concepts or semantic relation between these concepts. This paper focuses on the relation classification between relative concepts in scientific paper.

Relation classification is one of the most important topics for analyzing scientific papers. Most of the traditional relation classification methods

are influenced by the handcrafted features or extra NLP tools to derive lexical features (Surdeanu et al., 2012; Kozareva, 2012). However, these methods are time consuming and there is a problem of error propagation. Additionally, the traditional semantic textual similarity measuring approaches are using a large number of pairwise similarity features to represent the text. It is difficult for these features to represent the syntactic information.

To address these problems, DNN methods have been proposed and made remarkable achievement (Qin et al., 2016a; Guo et al., 2016). This paper is based on the work of (Qin et al., 2016b) which uses a CNN architecture to control feature learning automatically. As a result, Qin et al. (2016b) minimize the application of external toolkits and resources, which is used for part of speech (POS) or other basic pretreatment. Additionally, Zeng et al. (2014) proposes position feature to locate the entity pair, so as to highlight its promotion for the semantic relation. Owing to this position feature is mapped to several (e.g. 5) dimension followed each word's vector (e.g. 100 dimension), which represents the relative distances of current word to first and second entity. This position feature will disappear because of the excessive training times or error propagation during the training procedure. Thus we use the Qin et al. (2016b)'s entity tag features to strengthen the entity pair information, which use the tag words ($\langle e1s \rangle$, $\langle e1e \rangle$, $\langle e2s \rangle$, $\langle e2e \rangle$) to represent start and end position features of entities. What's more, these tag features are represented as independent vector so as to avoid position feature's disappeared defect in Zeng et al. (2014).

As far as we know, most of the pervious DNN methods used entire sentence's words embedding as the input information for DNN to extracting features for relation classification such as (Xu et al.,

*Corresponding author

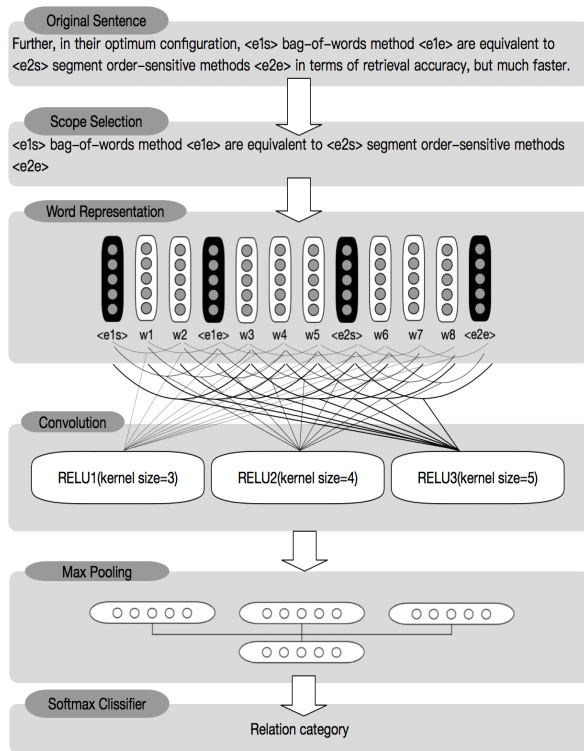


Figure 1: The architecture of CNN

2015; Liu et al., 2015). However, our goal is to achieve relation classification instead of sentence classification. Even though we use the position feature or entity tag feature to highlight the entity effect, it still suffers from that the long sentences have lots of noise words which is useless for relation classification. In pervious working, Qin et al. (2016b) just use the context between two entities, which got a remarkable performance promotion. Thus, we use Qin et al. (2016b)’s context scope as our CNN’s input information.

The contributions of this paper can be summarized as follows: Firstly, we construct a simple convolution neural network architecture for relation classification without sophisticated NLP preprocessing. Secondly, we use a more effective context input for convolution neural network, which extremely reduce the useless context’s noise effect. Then, we use entity tag feature to replace the entity position feature. Finally, we conduct experiments on subtasks 1.1 and 1.2 datasets, and the experiment results reveal that the proposed approaches is helpful to improve the performance.

2 Methodology

Our relation classification architecture is depicted as Figure 1. First, we select the scope of context

words and convert them to word embeddings. In the word representation step, the entity tag feature ($\langle e1s \rangle$, $\langle e1e \rangle$, $\langle e2s \rangle$, $\langle e2e \rangle$) will also be encoded into embeddings. Then, all the embeddings will be transmitted to three convolution network whose kernel size is 3, 4 and 5. Finally, these three convolution outputs are pooled into the same dimensional vector which will be concatenated as a input of a softmax classifier.

2.1 Context Scope for Convolution Neural Network

Most of the existing DNN relation classification methods use entire sentence’s words embedding as context information. As the following sentence, the entity *bag-of-words method* and *segment order-sensitive methods* have a **Compare** relation. However, the sub sentence *Further, in their optimum configuration and in terms of retrieval accuracy, but much faster* have little relevance to the target relation category. However, the sub sentence $\langle e1s \rangle$ *bag-of-words method* $\langle e1e \rangle$ are equivalent to $\langle e2s \rangle$ *segment order-sensitive methods* $\langle e2e \rangle$ have more relevant information to the target relation. As a result, we extract the context between two entities for relation classification. As for the no-words between entities’ condition, only two entities’ names have been extracted.

Further, in their optimum configuration, <e1s> bag-of-words method <e1e> are equivalent to <e2s> segment order-sensitive methods <e2e> in terms of retrieval accuracy, but much faster.

2.2 Convolution Neural Network Architecture

As the Convolution part in Figure 1, the input embedding is delivered to three convolution neural networks whose kernel size is 3, 4 and 5 respectively. It means that all the 3-grams, 4-grams and 5-grams features will be considered. Since each input sentence has a different length, the convolution output will be pooled into the same dimensional vector space. Finally, we use the multi-class classifier softmax to classify the relation into a specific category.

3 Experiment

3.1 Dataset

We use the SemEval-2018 Task 7.1.1 & 7.1.2 datasets Gábor et al. (2018) in experiment, which is in computational linguistic domain. This task

Embedding dim	Batch size	Dropout	Learning rate	Filter num	Epoch num	Activation
300	30	0.5	1e-4	100	30	Softmax

Table 1: CNN Parameters

Description	Macro-F1	
	sub1.1	sub1.2
CNN input: all words in the sentence (scope1)	47.2	67.8
+ Zeng et al. (2014)’s position feature	47.5	67.9
+Qin et al. (2016b)’s position feature	48.1	69.1
CNN input: entity pair and words between them (scope2)	48.2	68.9
+ Zeng et al. (2014)’s position feature	48.6	69.6
+Qin et al. (2016b)’s position feature	49.1	71.1
rules filtering before CNN	48.1	69.6
traditional pipeline method	42.3	54.6

Table 2: Experiment Results

addresses 6 semantic relation categories (5 relationships with bi-directions and an undirected *Compare* class) in scientific papers. We use datasets of subtasks 1.1 and 1.2 that contain titles and abstracts of papers in computational linguistic domain where entity mentions are either manually annotated (Subtask 1.1) or automatically annotated (Subtask 1.2). The relations are manually annotated and belong to 6 semantic relation categories (5 relationships with bi-directions and an undirected *Compare* class). There are 1228/1248 training examples and 355/255 testing examples in dataset 1.1/1.2. These relation instances are classified into one of the following relations: USAGE, RESULT, MODEL, PART-WHOLE, TOPIC, COMPARISON. The official evaluation metric is macro-F1 score.

3.2 Parameter Settings

The experiment settings are listed in Table 2.2, we use the Wikipedia general English 300 dimensional embeddings which have 408 million words¹. After testing, we find the parameters in the Table 2.2 achieve the most effective performance. As there are lots of parameters in CNN, we list some primary data in Table 2.2. For more detailed, we will share the whole project in our Github².

¹https://www.cs.york.ac.uk/nlp/extvec/wiki_extvec.gz

²https://github.com/zhongboyin/SemEval2018_task7

Rule	Category
”than” between entity pair	Compare
”used” between entity pair	Usage
”propose” between entity pair	Topic
”contain” between entity pair	Part-Whole
”precision” in entity name	Result
”model” in entity name	Model-Feature

Table 3: Filtering Rules.

3.3 Effect of Position Feature

As described in previous sections, position feature is helpful to promote the classification’s performance. Moreover, results in table 2.2 (first 3 lines or 4th to 6th lines) proved that Qin et al. (2016b)’s position features have a better performance than Zeng et al. (2014)’s features.

3.4 Effect of New Context Scope

By comparing 1st and 4th lines’ result, we could conclude that the entity names and words between them contain more accurate and cleaner semantic relation information.

By analyzing the predicted relation of the two experiments, we find that many wrong predicted long sentence instances in entire scope (scope1) experiment have been corrected in words between entity pair scope (scope2) as Table 4.

3.5 Result of Rule-based Experiment

In our early research, we explored lots of heuristic rules (as in Table 3.5) for each category by ob-

Sentence	Scope	Prediction	True/False
We present an implementation of the model based on finite-state models, demonstrate the $\langle e1s \rangle$ model’s $\langle e1e \rangle$ ability to significantly reduce $\langle e2s \rangle$ character and word error rate $\langle e2e \rangle$, and provide evaluation results involving extraction of translation.	scope1	Compare	False
	scope2	Result	True

Table 4: Input Scope Experiment Result Analyzing.

servicing the provided training set manually. However, most of sentences’ category couldn’t be determined by these rules. Thus, we divided this task into two-steps method. For the first step, we use the heuristic rules to classify some sentences into a specific category. Then, for the second step, we use our CNN method to classify the remaining sentence into a specific category. Before the testing set published, the heuristic rules achieved remarkable improvement in development set. On the contrary, the heuristic rules’ filtering step drops the performance in the testing set as the 7th line result in Table 2. After analyzing the result, we noticed that the rules are overfitting, since all of them are explored by training set and promoted by development set. As a result, it causes the decrease of the final evolution performance in the testing set.

3.6 Results of Comparison Experiment

To further prove the better performance of our CNN’s relation classification method, we also evaluated the same dataset using a more traditional NLP method which is based on the Multinomial Naive Bayes Classifier³. First the data from the training text file is extracted. The labels are extracted and encoded using a LabelEncoder. All the words from e1 to e2 in a sentence are considered for training the Multinomial Naive Bayes classifier. These words are also lemmatized and stemmed for better prediction. However, traditional pipeline method not only excises the error propagation problem, but also can’t detect some complicated semantic information such as hyponymy or synonymy. As a result, CNN method has a better performance than traditional method. The 8th line’s result in Table 2 shows that our CNN method is better than the general NLP processing method.

³<https://github.com/shouryay98/Hackabout-2017-Round-1>

4 Conclusion and Future Work

In this paper, we propose a new convolution neural network architecture for relation classification in scientific paper. We showed that the words between entity pairs are the most important for relation classification. Finally, our proposed method gets the macro-f1 value of 49.1 for subtask 1.1 and 71.1 for subtask 1.2.

For the future work, we will explore more features which are helpful for relation classification such as entity type and preposition features. Moreover, we will explore a more flexible sub-sentence scope as the context information for relation classification.

Acknowledgements

Firstly, we want to express gratitudes to the anonymous reviewers for their hard work and kind comments, which will further improve our work in the future. Additionally, this work was supported by the National Natural Science Foundation of China (No. 61602490).

References

- Kata Gábor, Davide Buscaldi, Anne-Kathrin Schumann, Behrang QasemiZadeh, Haïfa Zargayouna, and Thierry Charnois. 2018. Semeval-2018 Task 7: Semantic relation extraction and classification in scientific papers. In *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018)*, New Orleans, LA, USA.
- Jiang Guo, Wanxiang Che, Haifeng Wang, Ting Liu, and Jun Xu. 2016. A unified architecture for semantic role labeling and relation classification. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 1264–1274.
- Zornitsa Kozareva. 2012. Cause-effect relation learning. In *Workshop Proceedings of TextGraphs-7 on Graph-based Methods for Natural Language Processing*, pages 39–43. Association for Computational Linguistics.

- Yang Liu, Furu Wei, Sujian Li, Heng Ji, Ming Zhou, and Houfeng Wang. 2015. A dependency-based neural network for relation classification. *arXiv preprint arXiv:1507.04646*.
- Lianhui Qin, Zhisong Zhang, and Hai Zhao. 2016a. A stacking gated neural architecture for implicit discourse relation classification. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2263–2270.
- Pengda Qin, Weiran Xu, and Jun Guo. 2016b. An empirical convolutional neural network approach for semantic relation classification. *Neurocomputing*, 190:1–9.
- Mihai Surdeanu, Julie Tibshirani, Ramesh Nallapati, and Christopher D Manning. 2012. Multi-instance multi-label learning for relation extraction. In *Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning*, pages 455–465. Association for Computational Linguistics.
- Kun Xu, Yansong Feng, Songfang Huang, and Dongyan Zhao. 2015. Semantic relation classification via convolutional neural networks with simple negative sampling. *arXiv preprint arXiv:1506.07650*.
- Daojian Zeng, Kang Liu, Siwei Lai, Guangyou Zhou, and Jun Zhao. 2014. Relation classification via convolutional deep neural network. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 2335–2344.