Relation Extraction Toward Patent Domain Based on Keyword Strategy and Attention+BiLSTM Model (Short Paper)

Xueqiang Lv[1], Xiangru Lv[1], Xindong You[1](✉), Zhian Dong[1], and Junmei Han[2]

[1] Beijing Key Laboratory of Internet Culture and Digital Dissemination Research, Beijing Information Science and Technology University, Beijing, China
youxindong@bistu.edu.cn

[2] Laboratory of Complex Systems, Institute of Systems Engineering, AMS, PLA, Beijing, China

Abstract. Patent terminology relation extraction is of great significance to the construction of patent Knowledge graph. In order to solve the problem of long-distance dependency in traditional depth learning, a new method of patent terminology relation extraction is proposed, which combines attention mechanism and bi-directional LSTM model and with keyword strategy. Category keyword features in each sentence obtained by the improved TextRank with the patent text information vectorization added. BiLSTM neural work and attention mechanism are employed to extract the temporal information and sentence-level global feature information. Moreover, pooling layer is added to obtain the local features of the text. Finally, we fuse the global features and local features, and output the final classification results through the softmax classifier. The addition of category keywords improves the distinction of categories. Substantial experimental results demonstrate that the proposed model outperform the state-of-art neural model in patent terminology relation extraction.

Keywords: Patent terminology relation extraction · Patent knowledge graph · Keyword features · BiLSTM · Attention mechanism

1 Introduction

Automatic extraction of patent terminology relationship plays an important role in patent information retrieval, patent similarity detection, patent domain ontology construction, patent knowledge graph construction and latent semantic analysis.

In this paper, a new method of patent terminology relation extraction is proposed, which combines BiLSTM with Attention and keyword strategy and pooling layer are also added. The improved TextRank algorithm is used to extract the class keyword features, then the BiLSTM neural network and attention mechanism are used to extract the temporal information and sentence-level important information, then the key

X. Wang et al. (Eds.): CollaborateCom 2019, LNICST 292, pp. 408–416, 2019.
https://doi.org/10.1007/978-3-030-30146-0_28

features of each sentence are selected under the action of the pool layer, and finally the classification results are obtained through the full connection layer into the classifier. As a whole, the main contributions of this paper are as follows:

(1) *The class keyword features of each sentence are extracted by improved TextRank algorithm, which is combined with the patent information.*
(2) *The pooling layer is added parallel with the attention mechanism after the Bi-LSTM, which is fused as the input of Softmax classifier.*
(3) *Substantial experiments are conducted on different neural network model, which confirm the superiority of our proposed methods than the state-of-art neural network model.*

2 Related Work

At present, many scholars have done a lot of research on relationship extraction, and under the impetus of the actual needs, the relationship extraction technology has made great progress and has been widely used. Relationship extraction methods include pattern matching method, dictionary-driven method, statistics-based machine learning method and multi-method hybrid method [1]. Rink and Harabagiu [2] extracted semantic and lexical features from external corpus, and extracted semantic relations using Support Vector Machines (SVM) classifier. Zhang et al. [3] used Kernel method to extract entity relations, and discussed various Kernel methods to extract relations from free text.

In recent years, the use of depth learning method for entity relationship extraction has become the mainstream. In depth learning, we can automatically learn and acquire effective text features. This method achieves better performance than the traditional methods in many natural language processing tasks without using the basic natural language processing tools [4]. Liu et al. [5] proposed a new convolutional neural network, introduced a new coding method, which used the synonym lexicon to encode the input words and combined with lexical features to extract relations. Zeng et al. [6] proposed a convolution neural network method based on entity location information for entity relationship extraction. The problem of long-distance dependency is alleviated to some extent. Santos et al. [7] proposed a new pairwise ranking loss function in the task of using convolution neural networks to deal with relational classification. The network is categorized by rankings and achieved the best classification performance at the time. Zhang et al. [8] use RNN based on word position information to complete the task of relation extraction, which makes better use of the context information of entities. Zhou et al. [9] used Attention+BiLSTM model for relation extraction. After the BiLSTM model got the high-level semantics of sentences, Attention mechanism was used for high-level semantics representation, which improved the performance of relation extraction.

3 Patent Terminology Relation Extraction Model

This paper extracts patent terms based on BiLSTM. At first, the patent text is severed by commas, semicolons and periods, and the terminology features in each sentence are identified, and the location information is added, and the features of the category keywords in each sentence are obtained by the improved TextRank keyword extraction algorithm, and then the sentences and the extracted features are formed into a final vector matrix. Vector matrix is imported into BiLSTM model and attention probability is calculated by attention mechanism. The whole feature of text information is obtained to highlight the importance of specific words to the whole sentence. At the same time, the key feature of each sentence is selected as a local feature by using the maximum pooling layer. Finally, the whole feature and the local feature are fused, and the final classification result is output through the classifier. The complete framework of the model is shown in Fig. 1.

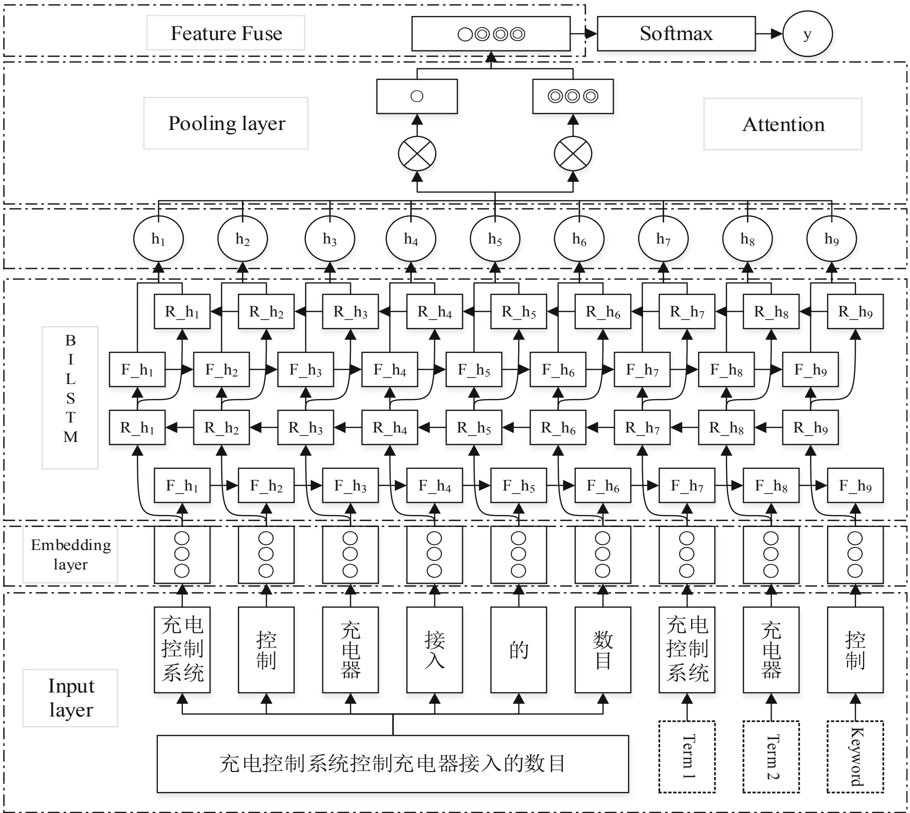

Fig. 1. Model complete framework

3.1 Position Vector Feature

In the task of patent term relation extraction, the words that can highlight the term relation are often distributed near the term, so adding the distance from each word to the two terms into the model can improve the effect of term relation extraction. For each word w_i in a sentence $\mathrm{s} = \{w_1, w_2, w_3, \ldots, w_k\}$ containing k words, the relative distance to the two terminologies is $i - i_1$ and $i - i_2$. Where i is the index of the position of the current word in the sentence, i_1 and i_2 are the index of the position of the two terms in the sentence, respectively.

3.2 Sentence-Level Category Keyword Feature Extraction

TextRank algorithm is simple and easy to use, which makes use of the relevance between words. But TextRank only relies on the document itself and the importance of each word is the same when initialized, so it is difficult to extract the keywords from the text accurately. TF-IDF algorithm relies on the corpus environment and can get the importance of a word in advance Therefore, combing the idea of TF-IDF algorithm to TextRank algorithm is helpful to improve the efficiency and accuracy of the algorithm. The improved Text-Rank (IMTR) algorithm is described as follows:

(1) Input patent text information set $\mathrm{S} = \{s_1, s_2, s_3, \ldots, s_n\}$, and set parameters: damping factor is d, sliding window size is w, maximum iteration number is I, iteration stop threshold is σ;
(2) The TF-IDF value of each word in the patent text information set S is calculated by a TF-IDF algorithm; And a keyword graph G_i composed of the words in s_i is constructed;
(3) According to the formula: $\mathrm{W}(v_i) = (1-d) + d \times W'(v_i)_{TF-IDF} \times \sum_{j \in In(v_i)} \frac{w_{ji}}{\sum_{v_k \in Out(v_j)} w_{jk}} W(v_j)$, the weight of each word in the keyword graph G_i is calculated iteratively until it converges;
(4) Each word in the keyword graph G_i is sorted by its weight, and the words with the largest weight and the verb part of speech are selected as the category characteristic keywords.

In the algorithm, $\mathrm{W}(v_i)$ is the weight of node v_i; d is damping factor; $In(v_i)$ is the set of nodes pointing to node v_i; $Out(v_j)$ is the set of nodes pointing from node v_j; w_{ji} is the weight of the edges of nodes v_j to v_i, and $W'(v_i)_{TF-IDF}$ is the TF-IDF value of node v_i.

3.3 Attention+BiLSTM Model

In the task of semantic relation extraction of patent terms, the historical information and future context information of the text should be taken into account. However, the LSTM model only records historical information and knows nothing about the future. Unlike the LSTM model, the bi-directional LSTM model considers both the characteristics of the past and those of the future. Simply understood, the bi-directional LSTM model is equivalent to two LSTMs, one forward output sequence and one reverse output sequence, and the outputs of the two are combined as the final result.

The bi-directional LSTM model effectively uses the context information of the patent text, and can extract more hidden features in the patent text.

In this part, we use the attention mechanism of relational classification task to calculate the output of Bi-LSTM model, and get the distribution of attention probability. From the distribution of attention probability, we can get the importance of the output state of LSTM unit to relational classification at each time, and then improve the final classification performance. In this model, the following formula is used for the attention layer:

$$\mathrm{M} = tanh(H) \tag{1}$$

$$\alpha = \mathrm{softmax}\left(\mathrm{w}^{\mathrm{T}}\mathrm{M}\right) \tag{2}$$

$$\mathrm{h}^{*} = \tanh\left(\mathrm{H}\alpha^{\mathrm{T}}\right) \tag{3}$$

Where $\mathrm{H} = [\mathrm{h}_1, \mathrm{h}_2, \mathrm{h}_3, \ldots, \mathrm{h}_{\mathrm{T}}]$ is a matrix output by that Bi-LSTM lay at T times and $\mathrm{H} \in \mathrm{R}^{\mathrm{d}^{\mathrm{w}} \times \mathrm{T}}$. d^{w} is the dimension of the word vector; w is the training parameter vector and w^{T} is the transpose of w; α is the probability distribution vector of attention; h^{*} is the expression of a learned sentence.

For the output $\mathrm{H} = [h_1, h_2, h_3, \ldots, h_T]$ of the BiLSTM model, besides the attention mechanism, the maximum pool method is used to compute the output, and the most relevant feature representation of the classification task is obtained, which is $h' = maxpool(\mathrm{H})$.

Feature fusion is to merge the computational results of attention layer and pooling layer to achieve the performance of complementary advantages among multiple features, which is $\mathrm{F} = h^{*} \otimes h'$. Where $\otimes$ represents vector splicing.

4 Experiment

4.1 Experimental Data and Evaluation Criteria

The data used in this experiment was a patent text of 9,978 new energy vehicles crawled from the patent search and analysis website. The ultimate goal of this experiment is to extract the terminology relation used in the patent text of the new energy automobile field. Since there are domain terms in each part of the patent text, the title, abstract, specification and claims in the patent are used as corpus. Patent text data were preprocessed and 6912 corpora were selected as experimental data, of which 5248 corpora were used as training data and 1664 corpora as test data. The data processing steps as follows:

(1) The patent terminology is extracted from the patent corpus by our previous proposed algorithm [10]. Dividing patent data into commas, semicolons and periods, each of which belongs to a corpus;
(2) Select a sentence that contains only two patent terms to form the final data set;
(3) Mark the selected data to determine the final experimental data.

There are 7 relationships in 6912 pieces of data selected in this experiment. The instance sample is shown in Table 1.

In order to verify the correctness and validity of the model proposed in this paper, the macro_averagedF1 (macro_F1) was used as experimental evaluation criteria. To calculate the macro-averaged F1 value, first calculate the Precision, Recall, and F1 value for each category. The formula is as follows:

$$P_i = \frac{TP_i}{TP_i + FP_i} \times 100\% \tag{4}$$

$$R_i = \frac{TP_i}{TP_i + FN_i} \times 100\% \tag{5}$$

$$F1_i = \frac{2 \times P_i \times R_i}{P_i + R_i} \times 100\% \tag{6}$$

TP_i is the number of data correctly predicted in the i-th relationship type. FP_i is the number of data erroneously predicted in the i-th relationship type. FN_i is the number of data belonging to the i-th relationship type that is incorrectly predicted to be of another relationship type. $macro_{averaged}F1$ is calculated as follows:

$$\text{macro\_averagedF1} = \frac{1}{M}\sum_{m=1}^{M} F1_m \tag{7}$$

Where M is the number of relationship types.

Table 1. Sample Instance.

Relation	Samples Content
Whole-Component	【驱动电机】装有两套【定子绕组】
Component-Whole	每组【动力电池】固定于一个【电池箱】中
Product-Material	其特征在于【转子】是用【稀土钴永磁】材料制作的
Spatial	【励磁绕组】与【电子控制器】相连接
Control	【操作手柄】控制【主轴】转动
Belongs to	所述【发动机】作为【动力单元】

4.2 Parameter Setting and Result Analysis

The experiments are conducted on a 64-bit Ubuntu 16.04 operating system installed on a Dell server with an NVIDIA Tesla K40 GPU and running memory of 64 GB. The model was implemented using the TensorFlow framework and python language. The experimental results of this model are closely related to the parameters in the model. Through a large number of parameter adjustment experiments, the local optimal value of each parameter is obtained. The Dimension of Word Vector is 300, the Dimension of Distance Vector is 50, the Batch_size is 128, the Learning Rate is 1e-5, the Hidden Layer is 256, the BiLSTM Layer is 2, the Droupout is 0.85. The overall results of this experiment are shown in Table 2.

Table 2. Overall experimental results.

Relation	Precision (%)	Recall (%)	F1 value (%)
Whole-component	95.97	93.53	94.73
Component-whole	87.55	95.61	91.40
Product-material	77.78	93.33	84.93
Control	97.06	83.90	90.00
Spatial	99.35	95.64	97.46
Belongs to	82.86	87.88	85.30
Other	95.38	84.62	89.68
Macro-averaging	90.85	90.64	90.50

From the experimental results for each relationship type in Table 2, from which we can be seen that the simplicity and complexity of the relationship types affect the final performance of the relationship extraction. This is because simple relationship types are easily learned by the model, and can be identified more accurately. It is difficult for the proposed model to learn the semantic association of complex relationship types, which result in low recognition accuracy.

4.3 Internal Comparison Experiment of the Model

In order to validate the effectiveness of keyword features and pooling layer adding Attention+BiLSTM model for patent terminology relationship recognition, four sets of internal comparison experiments are designed. The original input of the model is sentence vector, position vector and terminology vector. The experimental results are shown in Table 3.

Table 3. Comparative results of internal experiments of the model.

No.	Models	Evaluation criteria (%)		
		macro_P	macro_R	macro_F1
1	Attention+BiLSTM (ABL)	87.98	89.19	88.39
2	Keyword+Attention+BiLSTM (KABL)	89.50	89.63	89.34
3	Attention+BiLSTM+Pooling (ABLP)	88.77	89.32	88.81
4	Keyword+Attention+Bi-LSTM+Pooling (KABLP)	90.85	90.64	90.50

From the accuracy rate, recall rate and F1 value of each group of experiments shown in Table 3, we can see that the model of designed in this paper has got relatively good results and new energy vehicle patent terminology relation can be effectively extracted. In Experiment 1, only the Attention+Bi-LSTM model was used. Although performance has improved to some extend, the problem of terminology relation extraction in the patent field could be solved to a certain extent, but the final extraction result still needs to be improved. Experiment 2 added the keyword features on the basis of Experiment 1, and Experiment 3 added the pooling layer on the basis of Experiment 1. These two groups of experiments have improved experimental results compared to Experiment 1. It can be concluded that the keyword features and pooling layer have played a role in improving the efficiency of extraction of terminology relation in the patent domain. Compared to Experiment 1, the F1 value in Experiment 2 is increased by 0.95% and the F1 value in Experiment 3 increased by 0.42%. It can be concluded that the keyword features has played a greater role than the pooling layer in improving the efficiency of extraction of terminology relation in the patent domain. This is because the addition of category keyword features improves the distinction of categories of patent terminology relation, and also makes up for the shortage of Attention +BiLSTM model automatic learning features, therefore, the explicit addition of keyword features can play a certain role in patent terms relationship extraction.

Therefore, a method of adding keyword features and pooling layer to Attention +BiLSTM model is designed in this paper. It can be concluded from Experiment 4 that the KABLP can achieve a better performance than the general deep learning model.

4.4 Comparative Experiments of Different Classification Methods

In order to verify the advantages of Attention+BiLSTM model in patent terminology relation extraction, Attention+BiLSTM model is compared with RNN, LSTM and Bi-LSTM model on the same dataset. In order to unify the experimental standards, the input word vectors of all the models are the same, and the pooling layer is added to the models. The experimental results are shown in Table 4.

Table 4. Experimental results of different methods

NO.	Models	Evaluation criteria (%)		
		macro_P	macro_R	macro_F1
1	Attention+BiLSTM (ABL)	84.18	83.39	84.18
2	Keyword+Attention+BiLSTM (KABL)	86.24	88.96	87.54
3	Attention+BiLSTM+Pooling (ABLP)	87.18	89.19	88.12
4	**Keyword+Attention+Bi-LSTM+Pooling (KABLP)**	**90.85**	**90.64**	**90.50**

Comparisons of the different methods in Table 4 show that the BiLSTM method exhibits better performance than the LSTM and RNN methods. This is because the Bi-LSTM model not only considers the past characteristics but also the future characteristics, and effectively uses the context information of the patent text, which can extract more hidden features in the patent text. By adding Attention mechanism to Bi-LSTM model, the performance is further improved, because attention mechanism can highlight the importance of a particular word to the whole sentence by calculating the probability of attention, which can make the model pay more attention to the important information in patent text.

5 Conclusion and Future Work

In this paper, we mainly focus on relationship extraction from the new energy vehicle patent terminology, and propose an Attention+BiLSTM combined with keyword strategy and pooling layer of patent terminology relationship extraction method. However, this model can only extract the preset patent terms relationship types, how to extract the open domain relationship and automatically discover new patent terms relationship will be our main future work.

Acknowledgments. This work is supported by National Natural Science Foundation of China under Grants No. 61671070, National Science Key Lab Fund project 6142006190301, National Language Committee of China under Grants ZDI135-53, and Project of Three Dimension Energy Consumption Saving Strategies in Cloud Storage System in Promoting the Developing University Intension–Disciplinary Cluster No. 5211910940.

References

1. Rong, B., Fu, K., Huang, Y., Wang, Y.: Relation extraction based on multi-channel convolutional neural network. Appl. Res. Comput. **34**(03), 689–692 (2017)
2. Rink, B., Harabagiu, S.: UTD: classifying semantic relations by combining lexical and semantic resources. In: International Workshop on Semantic Evaluation, pp. 256–259. Association for Computational Linguistics (2010)
3. Zhang, X., Gao, Z., Zhu, M.: Kernel methods and its application in relation extraction. In: International Conference on Computer Science and Service System, pp. 1362–1365. IEEE (2011)
4. Collobert, R., Weston, J., Bottou, L., et al.: Natural language processing (almost) from scratch. J. Mach. Learn. Res. **12**(1), 2493–2537 (2011)
5. Liu, C.Y., Sun, W.B., Chao, W.H., et al.: Convolution neural network for relation extraction. In: Motoda, H., Wu, Z., Cao, L., Zaiane, O., Yao, M., Wang, W. (eds.) ADMA 2013. LNCS, vol. 8347, pp. 231–242. Springer, Berlin (2013)
6. Zeng, D., Liu, K., Lai, S., et al.: Relation classification via convolutional deep neural network. In: Proceedings of the 25th International Conference on Computational Linguistics, pp. 2335–2344 (2014)
7. Santos, C.N.D., Xiang, B., Zhou, B.: Classifying relations by ranking with convolutional neural networks. Comput. Sci. **86**(86), 132–137 (2015)
8. Zhang, D., Wang, D.: Relation classification via recurrent neural network [EB/OL]. https://arxiv.org/pdf/1508.01006.pdf. Accessed 05 Apr 2015
9. Zhou, P., Shi, W., Tian, J., et al.: Attention-based bidirectional long short-term memory networks for relation classification. In: Meeting of the Association for Computational Linguistics, pp. 207–212 (2016)
10. Lv, X.: Patent domain terminology extraction based on multi-feature fusion and BILSTM-CRF model. Front. Artif. Intell. Appl. **309**, 495–500 (2018)