2019 IEEE 5th International Conference on Computer and Communications

Inter-Personal Relation Extraction Model Based on Bidirectional GRU and Attention Mechanism

Yuming Li*†, Pin Ni*†, Gangmin Li*‡, Xutao Wang*, Zhenjin Dai*
*Research Lab for Knowledge and Wisdom, Xi'an Jiaotong-Liverpool University, China
†Department of Computer Science, University of Liverpool, UK
Email: ‡Gangmin.Li@xjtlu.edu.cn

***Abstract*—Inter-Personal Relationship Extraction is an important part of knowledge extraction and is also the fundamental work of constructing the knowledge graph of people's relationships. Compared with the traditional pattern recognition methods, the deep learning methods are more prominent in the relation extraction (RE) tasks. At present, the research of Chinese relation extraction technology is mainly based on the method of kernel function and Distant Supervision. In this paper, we propose a Chinese relation extraction model based on Bidirectional GRU network and Attention mechanism. Combining with the structural characteristics of the Chinese language, the input vector is input in the form of word vectors. Aiming at the problem of context memory, a Bidirectional GRU neural network is used to fuse the input vectors. The feature information of the word level is extracted from a sentence, and the sentence feature is extracted through the Attention mechanism of the word level. To verify the feasibility of this method, we use the distant supervision method to extract data from websites and compare it with existing relationship extraction methods. The experimental results show that Bi-directional GRU with Attention mechanism model can make full use of all the feature information of sentences, and the accuracy of Bi-directional GRU model is significantly higher than that of other neural network models without Attention mechanism.**

***Index Terms*—Relation Extraction, Bi-GRU, Attention**

I. Introduction

Extracting the relationship between characters from massive text data is an important research direction in the field of relationship extraction. It can be applied to many occasions such as search system, commercial promotion, and intelligence analysis, so it has been widely studied and developed. As a significant important branch of entity-relationship extraction, Inter-Personal Relationship Extraction (IPRE) aims to automatically extract social relationships between character entities from unstructured or semi-structured text data. IPRE is a practical and powerful information extraction technology, which is key for building a personal relationship network and has an important influence on the study of social networks. For example, in the text "Zhi Cao was the third son of Cao Cao and Empress Wuxuanbian", the character relationship corresponding to the person entities "Cao Cao" and "Zhi Cao" is "father-and-son", which can construct a triple "Cao Cao", "father-and-son relationship", "Zhi Cao", that laid the foundation for the subsequent knowledge graph construction.

This paper proposes a new BiGRU network that introduces the Attention mechanism to extract the relationship of the desired person entities in the text. The bidirectional context semantic information of the training instance can be obtained through the Bidirectional Gated Recurrent Unit (BiGRU), and the important semantic features in the instance are focused on by using the character level Attention mechanism. With the Attention mechanism in deep learning [1] has gradually played an important role in image classification, speech recognition, machine translation, etc. Some researchers have tried to use this method to reduce the impact of noise data. The Attention mechanism automatically focuses on words or instances that are critical to the RE, while ignoring the error labels generated by the distant supervisor building the training corpus to improve the performance of the relation extraction model.

The contributions of this paper can be listed as follows:

- Compared with LSTM, the representative varietal Recurrent Neural Network, the model of BiGRU is simpler, and it requires fewer parameters to meet the requirements of lightweight. Moreover, BiGRU neural network is less prone to over-fitting.
- Introduce Attention mechanism to capture the feature information of text sentences, and assign high weights to effective sentences through continuous learning, thereby reducing the negative impact of noise data;

II. Related Work

At present, the relationship extraction methods existing in academic circles can be roughly divided into two categories: supervised based traditional relationship extraction methods and deep learning-based relationship extraction. The traditional supervised relationship extraction system usually requires many manually labeled training data and automatically learns the corresponding extraction mode from the training data.

Supervised relation extraction methods mainly include: a method based on syntactic parsing enhancement [2], logistic regression-based method [3], kernel function-based method [4], [5] and Conditional Random Field-based method [6]. However, the shortcoming of these methods is that many traditional Natural Language Processing tools are used to extract advanced features such as part-of-speech tags, shortest dependent paths, and named entities, resulting in increased computational cost and additional error propagation. Another disadvantage is that owing to low coverage of entities and entity relationships of different training data sets, performances are not satisfiable in terms of versatility. In response to this limitation, Mintz et al. [7] proposed the idea of Distant Supervision, aligned relational entities in Freebase with Wikipedia articles and obtained a large number of relational instances as training corpus. The authors physically aligned the New York Times news text with the large-scale knowledge graph named Freebase, which contains more than 7,300 relationships and more than 900 million entities. The remote supervision assumes that a sentence containing two entities implies the relationship of the entity to Freebase and uses the sentence as a training example for the corresponding relationship of the entity. The author extracts the text features and trains the relational classification model on the data of the remote monitoring annotation, which effectively solves the problem of the size of the annotation data extracted by the relationship.

With the development of deep learning, deep learning algorithms have been gradually applied in relational extraction tasks since 2012. Socher et al. [8] proposed the use of recurrent neural networks to solve the problem of relation extraction. The method first parses the sentence and then learns the vector representation for each node on the syntax tree. Through the recurrent neural network, it can start from the word vector at the lowest end of the syntax tree, and iteratively merge according to the syntactic structure of the sentence, and finally obtain the vector representation of the sentence and use it for relationship classification. This method can take the syntactic structure information of the sentence into consideration effectively, but at the same time, it fails to consider the position and semantic information of the two entities in the sentence. Hoffmann et al. [9] and yang et al. [10] all adopted the strategy of multi-instance learning to reduce the influence of noise data, and achieved certain effects.

Zeng et al. [11] proposed the use of convolutional neural networks for relation extraction. They use the vocabulary vector and the position vector of the word as input to the convolutional neural network and get the sentence representation through the convolutional layer, the pooling layer, and the nonlinear layer. By considering the location vector of the entity and other related lexical features, the entity information in the sentence can be better evaluated in the relationship extraction. Later, Santos et al. [12] also proposed a new convolutional neural network for relation extraction, which uses a new loss function, which can effectively improve the discriminability between different relationship categories.

Under the condition of less manual intervention to get high-performance Relation Extraction model, based on the remote monitoring method of Relation Extraction [13] began to widely Attention this method under the assumption that if a relationship between two entities, then at least one contains the entity of the sentence describes the relationship between [14], under the premise of use of the existing knowledge base contains a relationship of entities, the text contains the entity to back to the sentence, with automatic access to a large number of training instances, can be better solve the problem of lack of labeled training data. Mintz et al. [7] aligned relational entities in Freebase with Wikipedia articles and obtained a large number of relational instances as training corpus. However, because the remote monitoring hypothesis is not strict, there is a large amount of noise data in the automatically constructed training corpus, which has a certain impact on the performance of the RE model. In order to solve this problem, various methods have been proposed. Hoffmann et al. [9] and yang et al. [10] all adopted the strategy of multi-instance learning to reduce the influence of noise data, and achieved certain effects. The methods in most of the above papers are oriented to English corpus, and only use the semantic information of words or instances in English. For Chinese Relation Extraction task, Chinese characters, as the most basic unit in Chinese, contain a lot of important semantic information, so the character-level information in Chinese training examples is important for Chinese Relation Extraction.

III. Methodology

A. IPRE Model

Our Inter-Personal Relationship Extraction Model can be divided into three layers: Embedding layer, BiGRU layer and Attention layer, the structure of the model can be shown in figure 1:

1) Embedding Layer: In the process of training the IPRE model with training examples, the first problem to be solved is to vectorize the text so that the model can be read. In this paper, we use Word2vec model to transform input Chinese characters into corresponding vectors. The model defines the input sentence as $S:\{x_1, x_2, ...x_n\}$ where n is the number of Chinese characters in the training example. For each Chinese character x_i in S, it is transformed into an embedding matrix $M^x \in R^{d*V}$ one by one, whereas d is the dimension of the word vector, V represents a fixed size vocabulary. Then through the matrix transformation according to $e_i = M^x \cdot V^i$ to converts x_i into the corresponding vector e_i, where V^i is a one-hot vector of V size. The output to complete the embedding is $E:\{e_1, e_2, ...e_n\}$.

2) BiGRU Layer: (Cho et al., 2014) proposed a Gated Recurrent Unit (GRU) that allows each recurrent unit to adaptively capture dependencies on different time scales. The activation value h_i of the hidden node at the current time of GRU unit can be expressed as:

$$h_i = z_i * h_{i-1} + z_i\widetilde{h_i} \tag{1}$$

where h_{i-1} represents the previous hidden state, z_i represent the update gate, and $\widetilde{h_i}$is the candidate value for the hidden node at the current time. The function of the update gate is to control how much historical information $h_{i-1}$is forgotten and how much current information $\widetilde{h_i}$ is remembered. The calculation of the update gate z_i is as follows:

$$z_i = \sigma(W_z e_i + V_z h_{i-1} + b_z) \tag{2}$$

where σ represents the sigmoid function, e_i is the i^{th} vector value output by the pre-training layer. while the candidate value $\widetilde{h_i}$ of the hidden node at the current time is calculated as:

$$\widetilde{h_i} = tanh(We_i + V(r_i * h_{i-1}) + b) \tag{3}$$

r_i and the reset gate

$$r_i = \sigma(W_r e_i + V_r h_{i-1} + b_r) \quad (4)$$

The smaller the reset gate, the less the information of the previous state is written. The output is $H : \{h_1, h_2, ...h_n\}$, $h_i \in R^2 S$ where S is the dimension of the hidden state in BiGRU By resetting gate and updating gate, GRU is able to learn long-distance information, which alleviates the problem of gradient disappearance or explosion caused by traditional recurrent neural network structure training. In addition, Bi-directional GRU can take into account both contextual and contextual information, which improves the global accuracy of relational extraction.

3) Attention Layer: The Attention mechanism was first applied in the field of images. There have been many experiments have shown that the accuracy of the recurrent neural networks with the Attention mechanism is significantly higher than that of the traditional neural network model. Since in the relationship extraction, the semantic information of each Chinese character contained in the relationship instance has different influence on the task, this paper introduces the character-level Attention mechanism to automatically pay Attention to the specific Chinese characters that play a decisive role in the relationship extraction and capture important semantic information in the instance. For the word vector matrix $[h_1, h_2, ..., h_n]$ output by the BiGRU network, where n is the number of Chinese characters contained in the relationship instance, by introducing a character-level Attention weight A for each weighted word vector. h_i:

$$W = tanh(H) \quad (5)$$

$$\alpha = softmax(\omega^T W) \quad (6)$$

$$A = tanh(H\alpha^T) \quad (7)$$

Where $H \in R^{d^w \cdot n}$, $W \in R^{d^w \cdot H}$, and d^w is the dimension of a word vector, w is a trained parameter vector.

Finally, after output through the Attention layer, it will be input into the Softmax layer for final relationship classification, hence, the final relationship result output can be determined.

B. Data sets and evaluation criteria

In our experiments, we used a static relationship extraction task dataset, which is collected from the crowdsourcing website, the Hudong Wiki (baike.com), using the remote supervision method to construct the relationship corpus. This corpus covers 12 types of relationship categories, which including ”parents”, ”couple”, ”teacher and student”, “brother or sister”, ”cooperation”, ”lovers”, ”grandparent and grandchild”, ”friend”, ”relative”, ”classmate”, ”superior and subordinate” and ”unknown”. The corpus is used as the experimental data as the basic resource for extraction. And the data structure of the train and the test datasets are ”entity1, entity2, relation type, description sentence”. We got 1000 data and divides it into a training set, validation set and a test set in the 8:1:1 ratio. For the experimental results, the accuracy (P), recall rate (R) and F1-measure (F1) are used as evaluation indicators.

The parameters of the model are set as shown in Table I.

IV. RESULT

In a word, we compared several algorithms that are efficient and popular in the current relationship extraction field. As shown in Figure 2, the accuracy of the GRU network is better than the LSTM network which both are the variant of the RNN model. The accuracy, recall and F1-Measure of the common BiGRU neural network model was 71.97%, 72.88% and 72.42%, respectively, and the BiGRU neural network model with the Attention mechanism was 82.67%, 80.08% and 81.35%, respectively. Regardless of the order of Accuracy, Recall, or F1-Measure, the model of the BiGRU neural network with Attention mechanism added is superior to the common BiGRU model, and it also has certain advantages compared to other methods. Therefore, it can be proved that the accuracy of relationship extraction can be improved by selectively focus on the important information in the sentence.

V. CONCLUSION AND FUTURE WORK

The Chinese character relationship extraction task has important research significance, and the results can be used to construct the Chinese character's knowledge graph, Chinese character relationship QA. Feature-based methods require manual definition of features and their extraction effects depend directly on the quality of feature selection. Kernel-based methods are slow and unsuitable for relational extraction tasks on large data sets. Moreover, it is difficult to extract relational patterns based on relational patterns, in turn, affects the extraction efficiency. Fortunately, the emergence of deep learning models solves these problems. This paper investigates and improves a deep learning-based method of relation extraction and the BiGRU-Att Chinese character relationship extraction model is proposed.

Future work may try to add a pre-training layer to the embedding layer to improve the accuracy of the model. At the same time, it will try to grasp the co-existing sentences of entities with the help of the search engine to construct training examples.

VI. ACKNOWLEDGEMENT

This work is partially supported by the AI University Research Centre (AI-URC) through XJTLU Key Programme Special Fund (KSF-P-02) and KSF-A-17. And it is also partially supported by Suzhou Science and Technology Programme Key Industrial Technology Innovation programme with project code SYG201840. We appreciate their support and guidance.

REFERENCES

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, and I. Polosukhin, "Attention is all you need," 2017.

[2] S. Miller, H. Fox, L. Ramshaw, and R. Weischedel, "A novel use of statistical parsing to extract information from text," in *North American Chapter of the Association for Computational Linguistics Conference*, 2000.

[3] N. Kambhatla, "Combining lexical, syntactic and semantic features with maximum entropy models for extracting relations," in *Meeting on Association for Computational Linguistics*, 2004.

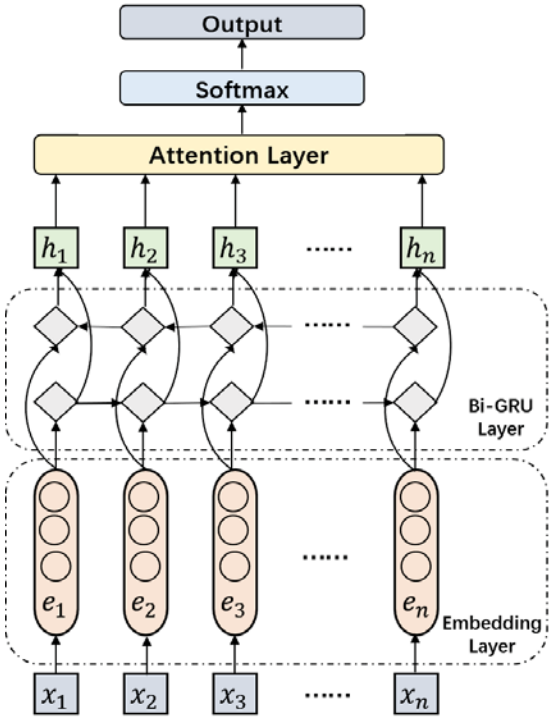

Fig. 1. The structure of the Inter-Personal Relationship Extraction Model

TABLE I PARAMETER SETTING

Parameter	Definition	Value
batch_size	Sample size for each training batch	64
epoch	The number of iteration	50
classes_num	The number of relationship categories set	11
gru_size	The number of GRU Network Unit	230
num_layer	The number of network layer	1
big_num	The number of entity pairs of each batch during training or testing	50

TABLE II COMPARISON OF EXPERIMENTAL RESULTS

Model	Accuracy	Recall	F1
BiLSTM [15]	70.13	76.07	72.98
BiGRU [16]	71.97	72.88	72.42
CNN [17]	71.91	76.87	74.31
Feature-SVM	69.57	60.71	64.84
LSTM-PCNN	77.35	70.11	73.55
LSTM-PCNN-Att	78.47	79.82	79.14
BiLSTM-CNN [18]	74.11	**80.13**	77.00
BiGRU-Att	**82.67**	80.08	**81.35**

Fig. 2. Visualization of BiGRU-Att Training Process

[4] S. Zhao and R. Grishman, "Extracting relations with integrated information using kernel methods," in *Proc Meeting on Association for Computational Linguistics*, 2005.

[5] R. C. Bunescu and R. J. Mooney, "Subsequence kernels for relation extraction," in *International Conference on Neural Information Processing Systems*, 2005.

[6] A. Culotta, A. Mccallum, and J. Betz, "Integrating probabilistic extraction models and data mining to discover relations and patterns in text," in *Main Conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, 2006.

[7] Mintz, Mike, Steven, Jurafsky, and Dan, "Distant supervision for relation extraction without labeled data," in *Joint Conference of the Meeting of the Acl the International Joint Conference on Natural Language Processing of the Afnlp: Volume*, 2009.

[8] R. Socher, B. Huval, C. D. Manning, and A. Y. Ng, "Semantic compositionality through recursive matrix-vector spaces," in *Joint Conference on Empirical Methods in Natural Language Processing Computational Natural Language Learning*, 2012.

[9] R. Hoffmann, C. Zhang, X. Ling, L. Zettlemoyer, and D. S. Weld, "Knowledge-based weak supervision for information extraction of overlapping relations," in *Meeting of the Association for Computational Linguistics: Human Language Technologies*, 2011.

[10] L. Yang, S. Liu, L. Kang, G. Zhou, and J. Zhao, "Exploring distinctive features in distant supervision for relation extraction," in *Asia Information Retrieval Symposium*, 2013.

[11] D. Zeng, S. Lai, X. Wang, L. Kang, J. Zhao, and X. Lv, "Distant supervision for relation extraction via sparse representation," 2014.

[12] C. N. D. Santos, B. Xiang, and B. Zhou, "Classifying relations by ranking with convolutional neural networks," *Computer Science*, vol. 86, no. 86, pp. 132–137, 2015.

[13] M. Craven and J. Kumlien, "Constructing biological knowledge bases by extracting information from text sources," 1999.

[14] S. Riedel, L. Yao, and A. Mccallum, "Modeling relations and their mentions without labeled text," in *European Conference on Machine Learning Knowledge Discovery in Databases*, 2010.

[15] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[16] A. Halidanmu, Y. Cheng, Y. Liu, and M. Sun, "Uyghur morphological segmentation with bidirectional gru neural networks," *Journal of Tsinghua University*, 2017.

[17] Y. Kim, "Convolutional neural networks for sentence classification," *Eprint Arxiv*, 2014.

[18] J. P. C. Chiu and E. Nichols, "Named entity recognition with bidirectional lstm-cnns," *Computer Science*, 2016.