

Research Article

Chinese Event Extraction Method Based on Roformer Model

Baohua Qiang ¹, Xiangyu Zhou,¹ Yufeng Wang,² Xianyi Yang ¹, Yuemeng Wang,²
Jubo Tian,² and Peng Chen¹

¹Guangxi Key Laboratory of Image and Graphic Intelligent Processing, Guilin University of Electronic Technology, Guilin 541004, China

²Hebei Key Laboratory of Intelligent Information Perception and Processing, The 54th Research Institute of CETC, Shijiazhuang 050081, China

Correspondence should be addressed to Xianyi Yang; xianyiyang65@126.com

Received 12 August 2022; Accepted 3 October 2022; Published 26 April 2023

Academic Editor: Xingsi Xue

Copyright © 2023 Baohua Qiang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Event extraction is an important research direction in the field of natural language processing. The current Chinese event extraction field still suffers from errors in the pretraining and fine-tuning stages, inability to directly handle texts with more than 512 tokens, and inaccurate event extraction due to insufficient semantic sample diversity. In this paper, we propose a Chinese event extraction method RoformerFC (Roformer model with FGM and CRF) based on the Roformer model to address the above problems. Firstly, our method utilizes the Roformer model based on rotary position embedding, which both moderates the errors in the pretraining and fine-tuning phases and allows the model to directly handle texts with more than 512 tokens; then, the adversarial networks based on FGM (fast gradient method) are realized to increase the diversity of semantic feature samples; finally, the classical CRF (conditional random fields) model is used to decode and identify the event element entity and its corresponding event role and event type. On the short text DuEE dataset, the microP, microR, and microF of our method improved 1.26%, 4.01%, and 2.68%, respectively, over the classical Chinese event extraction method BERT-CRF. On the long text JsEE dataset, the microP, microR, and microF of our method improved 2.26%, 5.03%, and 3.72%, respectively, over the classical Chinese event extraction method BERT-CRF.

1. Introduction

Natural language is a tool for human information exchange, and the connection between language and communication is inevitable. With the development of science and technology, natural language will get more applications in communication networks. Event extraction [1] technique for natural language processing is a key aspect of text processing, aiming at structuring free text and facilitating research in areas such as abstractive summarization [2] and information retrieval [3]. Event extraction is usually divided into trigger word recognition and event element recognition, and in recent years, trigger word recognition and event element recognition have been recognized as a whole to avoid the wrong transmission of the pipeline [4].

Compared with English, Chinese texts are characterized by semantic ambiguity, complex word boundaries, and high-dimensional sparsity [5]. Therefore, how to handle Chinese text is the focus of Chinese event extraction. Word embedding models or dynamic pretraining models are usually used to learn the contextual semantic features of Chinese text. The advantage of the word embedding model is that it is fast, but suffers from the problem of multiple meanings of words. In 2018, Google released the dynamic pretraining model BERT (Bidirectional Encoder Representation from Transformers) [6], which dynamically generates word vectors for input text, effectively eliminating the effect of polysemous words. However, the BERT model also has some problems, such as no mask sign appears in the fine-tuning phase, making the error between the pretraining and fine-tuning phases large and the

model cannot directly handle texts with more than 512 tokens. As the text length increases, the distribution of entities becomes wider. Using pretrained models to learn text semantics, a lack of semantic sample diversity occurs, resulting in low event extraction accuracy.

Since the long text event extraction dataset is not available in open source, we crawl nearly 10,000 military category news texts from major news websites such as Xinhua, People’s Daily Online, and NetEase in recent years and construct a Chinese long text event extraction dataset JsEE.

In the current event extraction field, there are still the following problems, such as the error in pretraining and fine-tuning stages. The pretraining model cannot process long text directly, and fewer semantic feature samples are generated according to the vocabulary. In this paper, a Chinese event extraction method RoformerFC (Roformer model with FGM and CRF) is proposed to address the above problem. Firstly, the method uses a Roformer model based on Rotary Position Embedding (RoPE) [7] to learn the contextual semantic features of Chinese text, which not only moderates the errors in the pretraining and fine-tuning stages but also allows the model to directly process texts with more than 512 tokens. Then a perturbation is added in the embedding layer using fast gradient method (FGM) [8] to increase the diversity of semantic feature samples and enhance the effect of event extraction. Finally decoding using conditional random field (CRF) identifies the event element entities and their corresponding event roles and event types.

The contributions of this study are as follows:

- (1) Using the Roformer model based on RoPE to learn semantic features of text, which both moderates the errors in the pretraining and fine-tuning phases and allows the model to directly process text with more than 512 tokens
- (2) Using the FGM approach to achieve adversarial training which means adding perturbations to the embedding layer to increase the diversity of semantic feature samples, thus improving the Chinese event extraction effect
- (3) Constructing a Chinese long text event extraction dataset JsEE

2. Related Work

The current event extraction task can be methodologically classified into pattern matching-based methods, feature extraction engineering-based methods, and deep learning-based methods [9]. Deep learning-based method is the dominant approach for event extraction in recent years, as it provides a more comprehensive representation of the raw data. The main works are found in Ding et al.’s study [10]. He proposed using bidirectional long short-term memory (BiLSTM) model for learning text feature information of semantics. Feng et al. [11] proposed a model incorporating RNN (recurrent neural network) and long short-term memory (LSTM) network for event extraction with good results; Chen et al. [12] combined CNN (convolutional neural network) and BiLSTM model to mine the hidden

relationship information between words. Although all the above methods alleviate the problem of gradient explosion in deep neural networks, they do not consider syntactic dependency information in sentences. In recent years, researchers have paid much attention to syntactic and semantic information, and the main works are in Gao et al.’s study [13]. He proposed a joint event extraction model, adding self-attention mechanism to BERT to achieve the extraction of events from TCM (traditional Chinese medicine) literature; Zhang et al. [14] proposed an end-to-end model based on BERT to improve the recognition accuracy of the elements belonging to each event by sequentially introducing the event types output from the antecedent layer and the entity embedding representation in the element and role recognition of the model; Chen et al. [15] proposed a financial event extraction method based on the pretrained model (Efficiently Learning an Encoder that Classifies Token Replacements Accurately) ELECTRA and lexical features, which enhances the perception of model, fully considering the original semantic information of the corpus as well as lexical feature information; Chen et al. [16] proposed a BERT-based event extraction model for news content, which improves the effect of event extraction by adding a DGCNN pooling layer. Xue and Huang [17] propose a generative adversarial network (GAN) with simulated annealing algorithm (SA-GAN), where the stagnation counter is introduced to accelerate GAN’s the convergence speed.

In summary, although there have been numerous methods for event extraction based on the BERT model since its release, most of them only take advantage of its learning capability without considering the errors in the pretraining and fine-tuning stages, the inability to directly handle texts over 512 tokens, and the lack of semantic diversity of samples.

3. Methodology

3.1. Model Framework. In this paper, we propose a Chinese event extraction method RoformerFC based on Roformer model. Firstly, the original corpus is divided according to word granularity; then it is input to Roformer layer, and the initial embedding of the text is obtained by word vectors, sentence vectors, and position vectors; then perturbations are added to the initial embedding by means of FGM to obtain the final embedding; finally, the CRF is used to obtain reasonable labeling results.

The framework of RoformerFC method proposed in this paper is shown in Figure 1. It mainly includes the Roformer layer, the FGM-style adversarial networks layer, and the CRF layer. The specific functions and implementation of each layer in the model will be described in detail in the next section.

3.2. Roformer Model. Roformer is a WoBERT model with absolute position encoding substitution for RoPE. WoBERT is a Chinese BERT model that continues to be trained using MLM pretraining tasks based on the open source RoBERTa [18] from HIT (Harbin Institute of Technology). RoBERTa uses dynamic masks—each time a sequence is entered into the model, a new mask pattern is generated. In the process of continuous input of a large amount of data, the model will gradually adapt to different mask strategies and learn different

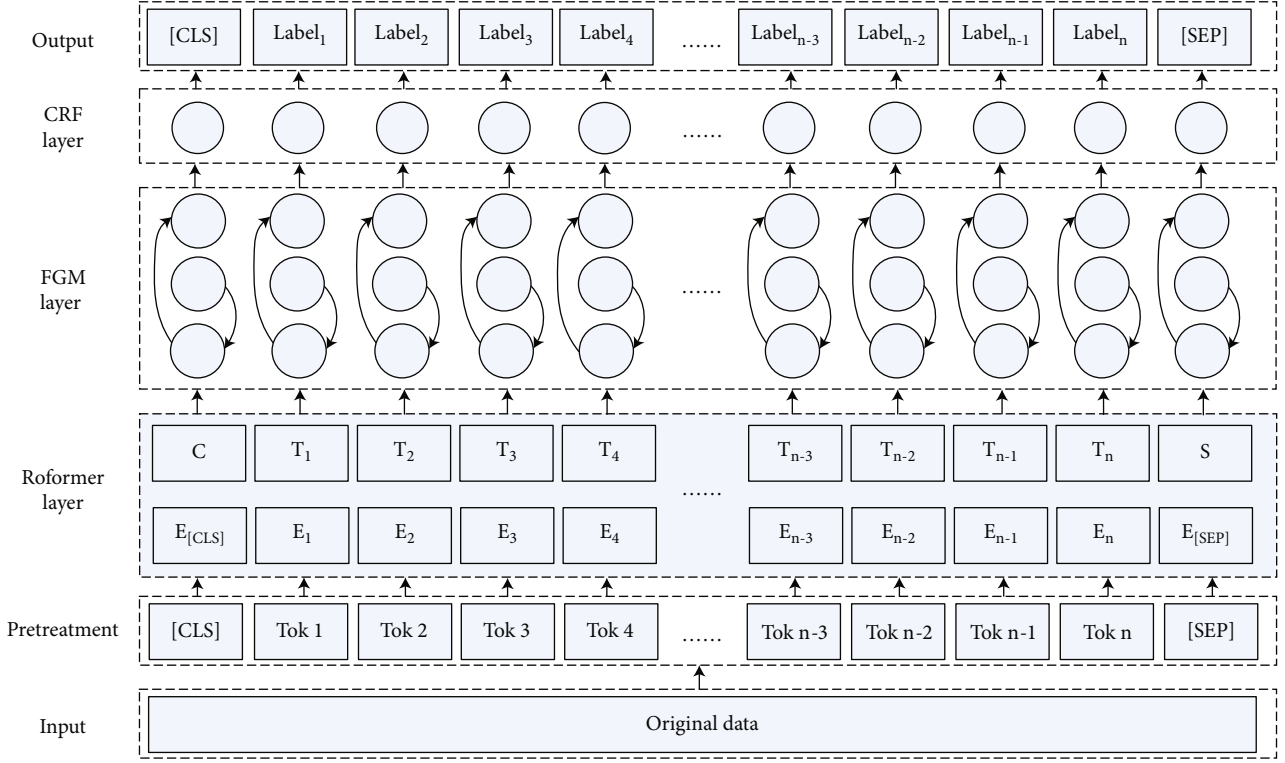


FIGURE 1: Schematic diagram of RoformerFC method framework.

language representations, thus moderates the errors in the pre-training and fine-tuning phases. The Roformer model relies on a new way of position encoding—RoPE.

The RoPE expressed as a complex number in the two-dimensional case is shown in the following:

$$f(q, m) = qe^{i(\Theta(q)+m\theta)}, \quad (1)$$

where q is the absolute position encoded vector, m is the location of the q vector, Θ is the complex number operation, $\theta = 10000^{-2/d}$, d is the dimensionality of the vector, and e is a constant in mathematics with a value of about 2.71828.

According to the geometric meaning of complex multiplication, it can also be written in matrix form, such as the following equation, which corresponds to the rotation of the vector by a certain angle, so it is called “rotary position embedding.”

$$f(q, m) = \begin{pmatrix} \cos m\theta & -\sin m\theta \\ \sin m\theta & \cos m\theta \end{pmatrix} \begin{pmatrix} q_0 \\ q_1 \end{pmatrix}, \quad (2)$$

where q_0 and q_1 are two-dimensional vector representations of the vector q . According to the operation rules of the matrix, any even-dimensional RoPE can be expressed as a splice of the two-dimensional case. The vector q is encoded in the absolute position at position m . The vector q is multiplied by the orthogonal matrix \mathcal{R}_m to obtain Q . The same operation is performed on the absolute position encoded vector k at position n to obtain K . By performing the attention operation on the obtained Q and K , the obtained sequence contains the relative

position information, as the constants establish the following:

$$(\mathcal{R}_m q)^T (\mathcal{R}_n k) = q^T \mathcal{R}_m^T \mathcal{R}_n k = q^T \mathcal{R}_{n-m} k. \quad (3)$$

Moreover, \mathcal{R}_m is an orthogonal matrix, and according to the paradigm preserving nature of orthogonal matrix, it does not change the length of the vector, but only rotates the original vector by a certain angle, so it does not change the internal structure and stability of the model.

3.3. Adversarial Training Based on FGM Approach. Adversarial training in deep learning and machine learning generally has two meanings. One is to generate adversarial networks to improve the learning ability of the model by increasing the difficulty of model training, and the other is to care about the robustness of the model under small perturbations. Due to the relatively small amount of data annotation, our method utilizes the FGM method to add perturbation to the embedding layer, and the learning ability of the model is improved by adding a certain number of negative samples. The input of CRF is obtained by summing the perturbation Δx and the word vector x after the sequence passes the Roformer model. The formula for implementing adversarial training based on the FGM approach is shown as

$$\min_{\theta} E_{(x,y) \sim T} \left[\max_{\Delta x \in \Omega} L(x + \Delta x, y; \theta) \right], \quad (4)$$

where T represents the training set for event extraction, x is the word vector input, y is the element label, θ is the hyper-parameter, $L(x + \Delta x, y; \theta)$ is the loss value of a single sample,

Δx is the adversarial perturbation added by means of FGM, and Ω is the perturbation space. For each sample, the perturbation Δx is added to x so that the larger the loss of individual samples, the greater the error in the model prediction as much as possible, thus improving the learning ability of the model.

Since the common way to reduce the loss is gradient descent algorithm, our method uses gradient ascent to calculate Δx . The reduction of interference loss enhances the training difficulty of the model, which makes the ability of model to learn deep semantic features of text increase, as shown in the following equation, where θ is the hyperparameter, x is the word vector input, y is the element label, $L(x, y; \theta)$ is the loss value of a single sample, Δx is the adversarial perturbation added by way of FGM, and ∇x denotes the gradient ascent calculation of x .

$$\Delta x = \nabla x L(x, y; \theta). \quad (5)$$

If Δx is too large, most of the semantic features learned by the model are wrong information, thus underfitting and reducing the effect of Chinese event extraction. To prevent Δx from being too large, the calculation is normalized as in the following equation, and the activation function is sign. sign means to divide those greater than 0 into 1 and those less than 0 into -1.

$$\Delta x = \text{sign}(\nabla x L(x, y; \theta)). \quad (6)$$

Equations (5) and (6) are substituted into Equation (4) to obtain the adversarial training formula based on the FGM approach, shown as

$$\min_{\theta} E_{(x,y) \sim T} [L(x + \Delta x, y; \theta)]. \quad (7)$$

3.4. CRF Model. In this paper, we use CRF to store the transfer probabilities between element entities and use the Viterbi algorithm [19] to find the shortest path for the transfer matrix to obtain the sequence with the highest probability. The probability value of the label sequence p_1, p_2, \dots, p_n is calculated using CRF for the input sequence vector q_1, q_2, \dots, q_n as in the following:

$$p(y|s) = \frac{\exp\left(\sum_i \left(W_{CRF}^{q_i} p_i + b_{CRF}^{(q_{i-1}, q_i)}\right)\right)}{\sum_{y'} \exp\left(\sum_i \left(W_{CRF}^{q'_i} p_i + b_{CRF}^{(q'_{i-1}, q'_i)}\right)\right)}, \quad (8)$$

where p_i is the input sequence, q_i is the sequence of element label, y' denotes any sequence of element label, $W_{CRF}^{q_i}$ denotes the hyperparameter of label q_i , $b_{CRF}^{(q_{i-1}, q_i)}$ denotes the bias of q_{i-1} from q_i , and $p(y|s)$ denotes the probability of the corresponding label of the sequence. During the training process, the maximum likelihood function of (p_i, q_i) is optimized by regularization, as shown in the following:

$$L = \sum_{i=1}^N \log(p(y_i|s_i)) + \frac{\lambda}{2} \|\theta^2\|, \quad (9)$$

where $p(y_i|s_i)$ is the probability value from the original sequence to the model predicted sequence, and λ and θ are regularization parameters.

3.5. RoformerFC Algorithm Implementation. The pseudocode of the RoformerFC algorithm is shown in Algorithm 1, which is applied to the short text DuEE dataset and the long text JsEE dataset, respectively. The input is the original free text and the hyperparameters to be used in the model training, and the output is the label of each word of the original free text. The set of label can be used to know which event element entities the text contains and their corresponding event roles and event types. There are three key modules in the algorithm: (i) the Roformer model based on rotational coding is used to enable the algorithm to directly process texts longer than 512 tokens, which is step 4; (ii) the FGM is used to dynamically add perturbations to the embedding matrix to improve the model robustness, which is step 6; (iii) the CRF is used to calculate the conditional probability of each character to improve the tag recognition accuracy, which is step 7.

4. Experiment

4.1. Dataset

4.1.1. Short Text DuEE Dataset. The Baidu Event Extraction Contest Dataset (DuEE) [20] is a public dataset released by Baidu PaddlePaddle AI Studio in April 2020 in the Dataset Hall, which was released to promote research in the field of human-computer interaction and natural language processing. DuEE contains 65 event categories with 213 event element categories, a total of 14945 data, and the training set, validation set, and test set are divided according to the ratio of 8:1:1. As shown in Table 1, 5 types of event types and their elemental framework are demonstrated.

By counting the 11,958 data in the training set, we found that 91.89% of the data were less than 200 characters in length, 61.06% were less than 100 characters in length, 29.83% were between 100 and 200 characters in length, less than 10% were more than 200 characters in length, and the data with more than 512 characters only account for 0.67%. In summary, the DuEE dataset is rich in event types, but the text length is mainly concentrated in the length of 100 characters or less.

4.1.2. Long Text JsEE Dataset. Through statistics, it is found that the data length in Baidu DuEE dataset is short, so we crawled nearly 10,000 military category news texts from major news websites such as Xinhua, People's Daily, and NetEase in recent years, organized and labeled the Chinese long text event extraction dataset JsEE. It includes 2 event types and 16 event element categories. The event types and event roles in this dataset are described below, along with the rules for annotation and the results of the annotation.

The dataset is labeled with 2 event types: Military-Multinational Joint Military Exercises and Military-Military Exercises. Each event type contains 8 event roles: Start Time, End Time, Country, Purpose, Exercise Name, Equipment, Military, and Location. Start Time means the start time of the military operation; End Time means the end time of the military operation; Country means the specific name of the country

Roformer algorithm

Input: Dataset D , short text DuEE dataset $N \subset D$, long text JsEE dataset $P \subset D$, hyperparameters: Learning rate, maximum text length (maxlen), training rounds (epochs), batch size (batch size), word vector dimension, CRF learning rate (crf_lr);

Output: Label set D' ;

- 1) Initialization: Hyperparameters in the model
- 2) For each
- 3) If the item= $1, 2, \dots, T_{\max}$ do
- 4) The original encoding of RoPE according to Eqs. (1) and (4) to obtain the encoding vector q_1, q_2, \dots, q_m that can directly handle n ($n > 512$) tokens
- 5) Randomly select the batch size data from the dataset D , obtain the longest character length m in the data, and encode the data into the Roformer model to obtain the vector matrix $E_{(batch_size, m, 600)}$
- 6) Adding a perturbation Δx to $E_{(batch_size, m, 600)}$ according to Eq. (7) to obtain $E'_{(batch_size, m, 600)}$ with negative samples
- 7) Calculate the conditional probability of each character according to Eq. (8)
- 8) End for
- 9) Until the maximum number of iterations T_{\max} is reached or the model converges
- 10) The shortest path is obtained by Viterbi algorithm, and the label set D' is output.

ALGORITHM 1: RoformerFC algorithm processing process.

TABLE 1: Short text DuEE dataset partial event framework table.

| Event type | Event element framework |
|-------------------------------|--|
| Product behavior-release | Trigger, time, release product, release party |
| Organizational relations-join | Trigger, time, joiner, organization joined |
| Disaster/accident-crash | Trigger, time, place, number of deaths, number of injuries |
| Finance/trading-listing | Trigger, time, place, listed company, financing amount |
| Finance/trading-sales | Trigger, time, seller, transaction, sale price, acquirer |

TABLE 2: Experimental environment configuration.

| Environment name | Experimental configuration |
|-------------------------|---|
| Operating system | Ubuntu 16.04LTS |
| CPU | Intel(R) Xeon(R) CPU E5-2603 v2 @ 1.80GHz |
| GPU | GeForce GTX 1080Ti |
| Python version | 3.6 |
| Deep learning framework | TensorFlow 1.14.0, Keras 2.3.1, Bert4Keras 0.11.1 |

participating in the military exercise or multinational joint military exercise; Purpose means the purpose of the military operation; Exercise Name means the code name for a military operation; Equipment means weapons used in military operations; Military means the name of a unit involved in military operations; Location means the location of the exercise in military operations.

Label according to the labeling rules of short text DuEE dataset, a total of 817 data were labeled, and the training set, validation set, and test set were divided according to the ratio of 8 : 1 : 1. By counting the length of the single data in the training set in the JsEE dataset, we found that the data are generally longer than the DuEE dataset, where the text length less than 100 characters accounts for only 0.15% of the data. Most of the text length is between 200 characters and 512 characters, occupying 69.08%, and the text length greater than 512 characters accounts for 19.72%.

TABLE 3: Model parameters.

| Parameters | Parameter value |
|-------------------------|-----------------|
| Learning rate | 0.00002 |
| Maximum text length | 600 |
| Batch size | 2 |
| Epoch | 20 |
| Perturbation parameters | 0.5 |

4.2. *Environment Configuration and Parameter Setting.* All experiments in this paper are conducted in the same experimental environment. The experimental environment configuration is shown in Table 2.

The pretrained models for all experiments in this paper use the base version and have the same hyperparameter settings, which are shown in Table 3.

TABLE 4: Experimental results. Unit: %.

| | microP | DuEE microR | microF | microP | JsEE microR | microF |
|--------------|--------|----------------|--------|--------|----------------|--------|
| BERT-CRF | 75.82 | 71.79 | 73.75 | 71.08 | 65.87 | 68.38 |
| NEZHA-CRF | 76.29 | 74.21 | 75.24 | 70.69 | 68.96 | 69.81 |
| Roformer-CRF | 76.06 | 75.62 | 75.84 | 71.26 | 69.52 | 70.38 |
| RoformerFC | 77.08 | 75.80 | 76.43 | 73.34 | 70.90 | 72.10 |

4.3. *Evaluation Criterion.* In the sequence labeling task of natural language processing, the evaluation metrics are mainly composed of three aspects, P value, R value, and F value. P value is the precision rate, R value is the recall rate, and F value is the comprehensive evaluation of precision rate and recall rate. In this paper, we improve on it by finding the average of multiple confusion matrices, which is called the microaverage judging metric, and the specific calculation process is shown in the following:

$$\text{microP} = \frac{\overline{\text{TP}}}{\overline{\text{TP}} + \overline{\text{FP}}}, \quad (10)$$

$$\text{microR} = \frac{\overline{\text{TP}}}{\overline{\text{TP}} + \overline{\text{FN}}}, \quad (11)$$

$$\text{microF} = \frac{2 \times \text{microP} \times \text{microR}}{\text{microP} + \text{microR}}, \quad (12)$$

where $\overline{\text{TP}}$ is the number of event element word levels for which the model correctly predicts both event types and event roles, $\overline{\text{FP}}$ is the number of event element word levels for which the model incorrectly predicts event types or event roles, and $\overline{\text{FN}}$ is the number of event element word levels for which the model does not extract the correct event element. microP is the microaverage precision rate, which refers to the proportion of correctly extracted event element entities to the extracted element entities in the event extraction results. microR is the microaverage recall, which refers to the proportion of correctly extracted event element entities in the event extraction results to the original manually annotated element entities. microF is the combined evaluation of microaverage precision rate and microaverage recall rate.

4.4. *Experimental Results and Quantitative Analysis.* In order to verify the effectiveness of the RoformerFC approach proposed in this paper, we conducted experiments comparing it with the following models on the short text DuEE dataset and the long text JsEE dataset, respectively.

- (1) *BERT-CRF [21] model.* BERT is used to learn deep semantic features of the text, and CRF is used to capture the dependency features between contextual tags
- (2) *NEZHA-CRF model.* NEZHA uses relative position encoding instead of absolute position encoding and can directly handle text with more than 512 tokens

- (3) *Roformer-CRF model.* Roformer employs rotational positional encoding, which means that it can directly process text with more than 512 tokens, while moderating errors in the pretraining and fine-tuning phases

The experimental results on the short text DuEE dataset and the long text JsEE dataset are shown in Table 4.

The following conclusions can be drawn from the experimental results in Table 4:

- (1) Comparing the performance of model BERT-CRF and model NEZHA-CRF in the experiments illustrates that the absolute position encoding of the BERT pretraining model cannot fully learn the text semantics when the text length is larger than 512 tokens. In contrast, NEZHA uses relative position encoding, which can handle texts with more than 512 tokens, and the NEZHA full-word mask strategy effectively mitigates the problem of large discrepancies between the pretraining and fine-tuning stages and improves the extraction effect
- (2) Comparing the performance of model NEZHA-CRF and model Roformer-CRF in the experiment shows that the Roformer model improves the extraction effect better than NEZHA using dynamic mask strategy and rotational position encoding
- (3) The RoformerFC method proposed in this paper is optimal in performance on two different length datasets. On the short text DuEE dataset, compared with the model Roformer-CRF, the microP , microR , and microF of our method were increased by 1.02%, 0.18%, and 0.59%, respectively; compared with model NEZHA-CRF, the microP , microR , and microF of our method were increased by 0.79%, 1.59%, and 1.19%, respectively; compared with the model BERT-CRF, the microP , microR , and microF of our method were increased by 1.26%, 4.01%, and 2.68%, respectively. On the long text JsEE dataset, compared to the model Roformer-CRF, the microP , microR , and microF of our method were increased by 2.08%, 1.38%, and 1.72%, respectively; compared with model NEZHA-CRF, the microP , microR , and microF of our method were increased by 2.65%, 1.94%, and 2.29%, respectively; compared with the model BERT-CRF, the microP , microR , and microF

of our method were increased by 2.26%, 5.03%, and 3.72%, respectively

Compared with the first three groups of experiments, we can draw a conclusion that compared with BERT and NEZHA, Roformer reduces the errors in the two stages of pretraining and fine-tuning and solves the problem that the pretraining model cannot directly handle long texts. Comparing the latter two groups of experiments, we can conclude that FGM type perturbation increases the diversity of semantic feature samples.

The experimental results fully corroborate the excellent performance of the RoformerFC method proposed in this paper in the Chinese event extraction task.

5. Conclusions

In this paper, we propose a RoformerFC Chinese event extraction method based on Roformer pretraining model and FGM-style adversarial training. Firstly, the method uses rotated positional encoding of Roformer as a word vector approach, which enables the model to directly learn deep-level features of text semantics with more than 512 tokens; secondly, perturbation is added to the embedding layer of the model by means of FGM to increase the sample diversity of the semantic features; finally, CRF is used to capture the dependency features among contextual tags, so as to infer the sequentially reasonable label. The comparative experiments on two different datasets show that the RoformerFC method proposed in this paper has better performance in Chinese event extraction, which verifies the effectiveness of using the Roformer model with rotational positional encoding to enhance the learning ability of text semantic features and improving the learning ability of the model through the FGM approach.

Data Availability

The data supporting the results of this study are public data sets. The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the Natural Science Foundation of Guangxi under Grant 2019GXNSFDA185006; in part by the Development Foundation of the 54th Research Institute of China Electronics Technology Group Corporation under Grant SKX212010053; in part by the by the Development Fund Project of Hebei Key Laboratory of Intelligent Information Perception and Processing under Grant SXX22138X002; in part by the Guilin Science and Technology Development Program under Grants 20190211-17 and 20210104-1; in part by the Innovation Project of GUET Graduate Education under Grants 2022YCXS061.

References

- [1] C. M. Ma, X. H. Li, Z. Li, H. R. Wang, and D. Yang, "Survey of event extraction," *Computer Applications*, pp. 1–20, 2022.
- [2] W. S. El-Kassas, C. R. Salama, A. A. Rafea, and H. K. Mohamed, "Automatic text summarization: a comprehensive survey," *Expert Systems with Applications*, vol. 165, article 113679, 2021.
- [3] S. Marcos-Pablos and F. J. García-Peñalvo, "Information retrieval methodology for aiding scientific database search," *Soft Computing*, vol. 24, no. 8, pp. 5551–5560, 2020.
- [4] T. H. Nguyen, K. Cho, and R. Grishman, "Joint event extraction via recurrent neural networks," in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 300–309, San Diego, California, 2016.
- [5] W. Xiang and B. Wang, "Survey of Chinese event extraction research," *Computer Technology and Development*, vol. 30, no. 2, pp. 1–6, 2020.
- [6] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "Bert: pre-training of deep bidirectional transformers for language understanding," 2018, arXiv preprint arXiv: 1810.04805.
- [7] J. Su, Y. Lu, S. Pan, B. Wen, and Y. Liu, "Roformer: enhanced transformer with rotary position embedding," 2021, arXiv preprint arXiv: 2104.09864.
- [8] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," 2014, arXiv preprint arXiv: 1412.6572.
- [9] L. Wang, R. X. Li, Y. H. Li, X. W. Gu, and Q. Yang, "Joint model for document-level event extraction without triggers," *Frontiers of Computer Science and Technology*, vol. 15, no. 12, pp. 2327–2334, 2021.
- [10] N. Ding, Z. Li, Z. Liu, H. Zheng, and Z. Lin, "Event detection with trigger-aware lattice neural network," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pp. 347–356, Hong Kong, China, 2019.
- [11] X. Feng, B. Qin, and T. Liu, "A language-independent neural network for event detection," *Science China Information Sciences*, vol. 61, no. 9, pp. 1–12, 2018.
- [12] B. Chen, Y. Zhou, and B. Liu, "Event trigger word extraction based on convolutional bidirectional long short term memory network," *Computer Engineering*, vol. 45, no. 1, pp. 153–158, 2019.
- [13] S. Gao, H. Tao, Y. Z. Jiang, Q. Jia, D. Z. Zhang, and Y. H. Xie, "Sentence-level joint event extraction of tradition Chinese medical literature," *Technology Intelligence Engineering*, vol. 7, no. 5, pp. 15–29, 2021.
- [14] H. Zhang, H. Song, S. Wang, and B. Xu, "A BERT-based end-to-end model for Chinese document-level event extraction," in *Proceedings of the 19th Chinese National Conference on Computational Linguistics*, pp. 390–401, Haikou, China, 2020.
- [15] X. Chen, L. Ni, and Z. Ni, "Extracting financial events with ELECTRA and part-of-speech," *Data Analysis and Knowledge Discovery*, vol. 5, no. 7, pp. 36–47, 2021.
- [16] A. Chen, Y. Ye, C. Wang, W. Wang, and B. Li, "Research on Chinese event extraction method based on BERT-DGCNN," *Computer Science and Application*, vol. 11, no. 5, pp. 1572–1578, 2021.

- [17] X. S. Xue and Q. H. Huang, "Generative adversarial learning for optimizing ontology alignment," *Expert Systems*, vol. 39, pp. 1–12, 2022.
- [18] Y. Liu, M. Ott, N. Goyal et al., "Roberta: a robustly optimized bert pretraining approach," 2019, arXiv pre-print arXiv:1907.11692.
- [19] C. Park, Y. Jung, J. Kim, and Y. Jung, "Joint viterbi detection and decoding algorithm for bluetooth low energy systems," *Electronics Letters*, vol. 56, no. 6, pp. 310–312, 2020.
- [20] X. Li, F. Li, L. Pan et al., "DuEE: a large-scale dataset for Chinese event extraction in real-world scenarios," in *Natural Language Processing and Chinese Computing*, pp. 534–545, Springer, Cham, 2020.
- [21] P. Li, "Prosodic unit boundary prediction of Myanmar based on BERT-CRF model," *Computer Science and Application*, vol. 11, no. 3, pp. 505–514, 2021.