MPTTF-BERT: Multi-Prefix-Tuning Template Fusion By BERT For Zero-shot English Text Relation Extraction Model

Fangqi Song*

School of Foreign Languages, Zhengzhou University of Science and Technology, Zhengzhou 450064, China.

**Corresponding author. E-mail: ancrum@qq.com*

Received: Jan. 09, 2025; Accepted: Mar. 06, 2025

Relation extraction is the core task of information extraction. How to extract important relation features quickly and accurately from massive English texts has become a difficult task in English text relation extraction. Meanwhile, Prefix-tuning is widely used in zero-shot natural language processing tasks, but the existing zero-shot relationship extraction model based on Prefix-tuning is difficult to construct answer space mapping and depends on manual template selection, which cannot achieve good results. To solve the above problems, this paper proposes a novel zero-shot English text relation extraction model based on multi-Prefix-tuning template fusion via BERT. Firstly, the task of extracting zero-shot relation is defined as the task of mask language model, and the construction of answer space mapping is abandoned. The words output from template and the relation description text are compared in the word vector space to judge the relation category. Secondly, the part of speech of the description text of the relation class to be extracted is introduced as the feature, and the weight between the feature and each template is learned. Finally, this weight is used to fuse the results of multiple template outputs to reduce the performance penalty caused by manually selected Prefix-tuning templates. The experimental results show that the F1 values obtained by MPTTF-BERT model on DuIE, COAE-2016-Task3, FinRE data sets with different features are 93.73%, 91.49% and 49.46%, respectively, which is significantly better than the comparison models. In addition, the ablation experiment and fixed-length selection experiments are used to further verify that the MPTTF-BERT model can effectively improve the effect of English text relation extraction, indicating the feasibility and effectiveness of the new method.

Keywords: zero-shot English text relation extraction; multi-Prefix-tuning template fusion; BERT; mask language model.

http://dx.doi.org/10.6180/jase.202512_28(12).0005

1. Introduction

Information extraction is an important technology and research field to extract useful information quickly and accurately from massive text data, which is of great significance to the construction of knowledge base in related fields. Relationship extraction is a key link in information extraction tasks. Its main purpose is to automatically extract semantic relations between subject and object on the basis of entity recognition, so as to complete the construction of entity relation triples < head entity, relationship, tail entity> [1–3]. Here,, the head entity and the tail entity are subject and object respectively, and relationship refers to the description of the relationship between the subject and object. The study of relation extraction technology is of great value to information acquisition [4], automatic construction of knowledge base [5], question and answer retrieval [6] and other related fields [7, 8].

In the real world, people often use short sentences to express and communicate, resulting in more and more short texts. Because short texts contain less semantic information, and from the linguistic perspective, English has more special and complex grammatical structures than other languages, English short texts have problems such as fewer features and more noise [9]. How to quickly process and extract important relational features from English short texts has become the key task of English short text relational extraction.

Early relation extraction tasks are mainly based on rules, eigenvector [10] and kernel function [11]. Qin et al. [12] completed the task of English entity relation extraction by introducing vocabulary, semantic matching and association rules. Geng et al. [13] enriched the feature expression among entities by integrating features such as semantic role annotation, morphology, and dependency syntactic relations, thereby improving the effect of relation extraction. Muhammad et al. [14] used semantic sequence kernel function combined with K-nearest neighbor algorithm and improved radial basis kernel function combined with polynomial function and convolution tree kernel function to effectively extract English text.

In recent years, deep learning methods have been widely used in relation extraction tasks, and have achieved better extraction results than previous methods. Based on the data set COAE-2016-Task3, Wang et al. [15] used convolutional neural network (CNN) to extract English entity relationships, and the accuracy rate, recall rate and F1 value in the best case reached 60.83%, 55.60% and 56.69% respectively. Considering the sequential nature of text data, it is proposed to apply recurrent neural network (RNN) to relational extraction tasks. Miwa and Bansal [16] used RNN method for the first time to complete the task of relation extraction, and used recurrent neural networks to parse the relation text, thus effectively taking into account the sequence structure of the text and improving the effect of relation extraction. In order to solve the problem of gradient disappearance and explosion caused by RNN, people begin to use the variant neural network of RNN and the hybrid method of multiple neural networks. Geng et al. [17] proposed a Long and Short Term Memory (LSTM) neural network model for relationship extraction, which showed good relationship extraction effect on public data sets. Chen and Hu [18] utilized Bidirectional Long Short Term memory networks Memory (BiLSTM) and conditional random field (CRF) to extract open domain relation, and the effect was significantly improved, which verified the effectiveness of BiLSTM in relation extraction. Xu et al. [19] proposed a hybrid model combining CNN and RNN to extract biomedical relationships. However, most of the above methods can only be used as a general framework, which is not applicable to a specific field. In order to solve this problem, attention mechanism is introduced on the basis of the general model, and different weights are assigned to different feature information to improve the effect of relationship extraction. For example, Hou et al. [20] proposed a CNN model integrating multiple attention mechanisms to improve text classification. Nayak and Ng [21] used the attention mechanism to explore the contribution degree of different words to entity relations and the potential semantic information to obtain rich context information, thereby improving the accuracy of relation extraction. Ahmad et al. [22] used the BiLSTM network model integrating multi-feature self-attention to predict the relationship. Cabot and Navigli [23] proposed CNN+Attention model to extract Chinese character relationships, and the F1 value obtained in the experiment reached 94.49%, which had a good extraction effect. Along with the proposed Transformer model [24], pre-trained models such as BERT are also used in relational extraction tasks. Han and Wang [25] proposed a text classification model based on BERT. Christou and Tsoumakas [26] built a BERT-GRU-ATT model and embedded relational texts through pre-trained model BERT, thus completing the classification of English entity relations. However, the pre-trained model usually needed to load many parameters, which leaded to the time cost of the model to complete the training, and had certain requirements for the hardware configuration.

Chen and Li [27] put forward the concept of zero-shot learning (ZSL), its core idea was to hope that the computer could simulate the way of human reasoning, and then identify new things that had never been seen before. In a general supervised learning task, the classes of the test phase must exist in the training phase, that is, all the classes are visible. In the task of ZSL, there are visible classes and invisible classes in the training and testing stages respectively. Through the learning in the training stage, it is necessary to identify the samples of invisible classes in the testing stage. In order to make the model predict the invisible class, the existing research focuses on the modeling process of the task, and the usual method is to design the zero-sample RE task into different task forms. For example, Caufield et al. [28] designed the task in the form of question and answer. Liu et al. [29] designed the task as a text implication problem. However, this method could not form an effective relational semantic representation space, and there was a large gap between tasks, and the performance of the model was usually poor. In recent years, pre-trained language models have alleviated the problem of inadequate representation of text semantic space, and the focus of ZSL task research in NLP field has gradually shifted to better use of pre-trained language models.

The pre-trained language model represented by the Transformer based on bidirectional encoder representations from Transformer (BERT) [30] has brought the NLP field into a new stage of development, and NLP tasks have begun to adopt the paradigm of fine-tuning the pre-trained language model in downstream tasks. Prefix-tuning is a technique to stimulate the hidden knowledge required by the pre-trained model to handle downstream tasks by giving prompts. The pre-trained language model can be deployed on specific tasks by reducing the gap between the pre-training phase and the downstream task phase by converting the original task into the training task of the pre-trained model. Chen et al. [31] tested the knowledge contained in the pre-trained language model by using cloze forms, proving that the pre-trained model could effectively preserve factual knowledge. However, at present, the construction of Prefix-tuning template in zero-shot environment is mostly manual, which is time-consuming and laborious. Zhao et al. [32] showed that the selection of Prefix-tuning template was counter-intuitive. In addition, Zhao et al. [33] proposed that the answer mapping obtained by manual design and gradient descent would bring about problems of high bias and high variance caused by incomplete coverage. To sum up, the efficient application of the traditional Prefix-tuning paradigm on the zero-shot RE task has two problems: the dependence of manual selection and the difficulty of constructing an answer mapping.

To solve these problems, this paper presents a zero-shot RE model based on multi-prefix-tuning template. The model transforms the zero-shot RE task into a relational representation generation task, directly abandons the traditional answer space mapping, and aligns the word vector space and the relational representation space. Through comparing the similarity between the word vector [MASK] output by the pre-trained language model and the word vector of the relational description text, it can determine the relation category to which it belongs. In addition, in view of the large differences in the representation space generated by different templates and the dependence of manual selection on template selection, this paper proposes a multi-prefix-tuning template fusion method, which assigns different Prefix-tuning templates weights according to the parts of speech of the relationship description text. By these weights, multiple Prefix-tuning templates are fused to improve the model RE capability.

2. Materials and methods

2.1. Problem definition and solution

The RE task is generally defined as: identifying the semantic relationship between s and o in x given the input sentence, subject and object (x, s, o). Where $x = \{x_1, x_2, \ldots, x_1\}$ represents the text of the sentence, l represents the length of the sentence, and x_i represents the $i-th$ token in the sentence. s and o represent the subject and object of the pair of known entities in the sentence. Further, the task of zero-shot RE is defined as: identifying the relation class r between the specified entity pairs of the sentence in the relation class set $R = R_{\text{train}} \cup R_{\text{test}}$. The class of relation to be identified in the test set R_{train} does not exist in the training set R_{test}, that is, the class of relation in the training set and the class of relation in the test set do not intersect $R_{\text{train}} \cap R_{\text{test}} = \emptyset$. In addition, each relation in R has a corresponding text description $D = \left\{d_1, d_2, \ldots, d_{|R|}\right\}$, where $|R|$ is the number of categories in the set of relation categories. For example, in the FewRel dataset, the relation id "P177" is "crosses" and the relation is described as "obstacle(body of water, road,...). which this bridge crosses over or..." .

Obviously, the supervised learning paradigm on the traditional RE task cannot be directly transferred to the zero-shot RE task. In this paper, the Prefix-tuning paradigm is introduced to transform the zero-shot RE task into a relational representation generation task. In this task, it is assumed that there exists a model M that can output a sentence and a relational representation of the correspondence between subject and object in the sentence. Formally, for the input sentence instance and the subject s and object o in the sentence, model $\mathrm{M}_{\mathrm{T_n}}$ outputs the current potential relationship representation, as shown in equation (1):

$$\mathrm{M}_{\mathrm{T_n}}(\mathrm{x},\ \mathrm{s}, \mathrm{o}) \rightarrow \tilde{\mathrm{r}}. \tag{1}$$

Where $\tilde{r} \in R^{\beta \times d}$ represents the generated relation. β is the number of predefined relational representations, i.e., the number of [masks]. d is the dimension of relational representation, which is consistent with the word vector dimension in the pre-trained language model. $\mathrm{T_n}$ indicates the current template. By measuring the distance between this representation and the representation produced by each candidate relation description text, the most appropriate category for the current text semantics is selected. Specifically, the Prefix-tuning template is embedded in the input, and the pre-trained language model will output the word vector of the [MASK] position in the template as a relationship representation; Then, the different relationship representations generated by different templates are fused to generate the final relationship representation.

2.2. RE model based on single Prefix-tuning template

The core purpose of zero-shot RE is to link sentences, subjects and objects with their corresponding relationship descriptions. Because the relationship of the test set does not exist in the training set, it usually requires a large amount of data or complex models to obtain the representation ability of describing the text. However, the task characteristic of zero-shot cannot provide sufficient data to support model training effectively. To solve the above problems, the direct use of the word vector space of the language model can ensure that the model can establish a better semantic representation under limited training samples. In this paper, the Prefix-tuning paradigm is used to generate the representation of the relationship, so as to realize the link between the sentence, the subject and the object and the corresponding relationship description.

This paper converts each input (x, s, o) to a new token sequence by filling it in with a predefined template. Each token represents a text unit or a supplementary symbol such as [MASK], and the new input is constructed as shown in equation (2):

$$\tilde{x} = [\mathrm{CLS}]x[\mathrm{SEP}]\mathrm{T_n}(s, o)[\mathrm{SEP}]. \tag{2}$$

The special characters set [CLS], [SEP] in pre-trained language model BERT that have no practical meaning, indicating the end of a sentence and the end of an entire paragraph of text, respectively. Formula (2) represents the combination of input sentences and template $T_n(s, o)$:

$$\begin{aligned}\mathrm{T_n}(s, o) = &[t_0 t_1 \dots t_c]\, s\, [t_{c+1} t_{c+2} \dots t_e] \\ &o \left[t_{e+1} t_{e+2} \dots t_{|\mathrm{T_n}(s,o)|-1}\right]\end{aligned} \tag{3}$$

Where $T_n(s, o)$ is the template selected from the common template library. t_{num} is a word in the template that satisfies $0 \leq \mathrm{num} \leq \left|T_{n(s,o)}\right| - 1$. $\left|T_{n(s,o)}\right|$ indicates the length of the template, $0 \leq c < e < \left|\mathrm{T}_{n(s,o)}\right|$. The pre-trained language model outputs a word vector for the [MASK] position in the template, and in this way obtains a relational representation.

Typically in Prefix-tuning paradigm, model $\mathrm{M}_{\mathrm{T_n}}$ predicts the $\mathrm{Probability_M}([\mathrm{MASK}] = v \mid \tilde{x})$ of the token being filled to the [MASK] position. In order to map the probability of v to the relational label, it needs to define the answer space map f: $v \mapsto \tilde{r}$, so that the model maps the token of the prediction. Li et al. [34] applied the Prefix-tuning paradigm to tasks such as emotion classification in this way. Unlike the simple and straightforward mapping between the prediction words "good" and "Positive" in these tasks, the RE task requires the extraction of complex semantic information between entities, including "crosses" and "original language of film or TV show" and so on, it is difficult to correspond them to specific words, that is, it is difficult to construct an answer map. Also, mappings that work well for visible classes may not work well for invisible classes, so additional validation data is often required to update templates and mappings. This approach is obviously not suitable for zero-shot RE tasks.

It is worth noting that in the zero-shot RE task, the corresponding description of the relation is the token sequence, and the word generated by the cloze can be directly compared with the relation description in the word vector space. For example, when entering a sentence instance from the relational category "P177" into the pre-trained language model in Prefix-tuning mode, "[CLS]...Yellow River bridge to replace the deteriorating Cape Girardeau Bridge. [SEP] The cape Girardeau bridge [MASK] the Mississippi River. [SEP]", the word that the model outputs to fill the covered position is "crosses", which exactly coincides with the name of the label relation class, indicating that word vector Spaces that directly use relation to describe text can also serve as class representation spaces. To sum up, in order to minimize the difference between the RE downstream task and the pre-trained language model, this paper disregards the constructed answer space mapping, and directly uses the generated word vector as a relational representation, and compares this representation with the word vector generated by the relational description after the pre-trained language model. As shown in equation (4), the word vector generated by MLM is compared with the word vector of the relational description text using Euclidean distance:

$$f_{\mathrm{dis}}(\tilde{r}, q) = \min_{i \in \{0,1,2,\dots,N\}} \sqrt{\sum_{j=1}^{d} (\tilde{r}_j - q_{ij})^2}. \tag{4}$$

Where $q \in R^{N \times d}$ is the word vector generated by the pre-trained language model to describe the text. N is the length of the description text. d is the dimension of the word vector.

Table 1. Meanings of some parts of speech and their abbreviations in NLTK library

Part of speech abbreviation	Explanation
NN	noun, singular
IN	preposition/ subordinating conjunction
KK	adjective
VBN	verb, pastparticiple
NNS	nounplural
DT	determiner
TO	to
VBG	verb, gerund/ present participle taking

Table 2. Experimental information for the dataset

Data set	Relation class	training set	verification set	testing set	Maximum length	text
DulE	49	62280	7793	7849	160	
COAE-2016-Task3	10	790	198	483	129	
FinRE	44	13487	1489	3720	97	

Multi-Prefix-tuning template fusion method based on part-of-speech

The experiment finds that after the same example is input into different Prefix-tuning templates, there are differences in the output of the model, which is generally reflected in the ability of the model to extract different relationships. The selection of Prefix-tuning template plays a very key role in the performance of the model. However, the existing automatic template selection algorithm requires a certain amount of data and is not suitable for zero-shot task. Therefore, this paper presents a multi-prefix-tuning template fusion method suitable for zero-shot task. As shown in Figure 1, the output results of multiple templates are fused by introducing parts-of-speech information common to both visible and invisible classes.

In the training phase, the weights of different Prefix-tuning templates are learned according to the parts of speech of the relational description text, and multiple templates are fused according to these weights in the test phase. Specifically, NLTK (Natural Language ToolKit) library is used to obtain the parts of speech of the relational description text and delete the stop words in the text. The feature vector $\mathrm{H_i} = \{h_{\mathrm{NN}}, h_{1\,\mathrm{N}}, \ldots, h_{\mathrm{KK}}\}\ (\mathrm{i} \in \{1, 2, \ldots, |\mathrm{R}|\})$ is assigned to each relation according to the part of speech feature, where $H_i \in R^{\gamma \times 1}$, γ is the class number of the part of speech tag, and $|R|$ is the class number of the relation class. $h_{\mathrm{NN}} = \text{frequency } \mathrm{N_i}/\gamma$ is an eigenvalue in the description text where the part of speech is NN. frequency $_i^{\mathrm{NN}}$ indicates the frequency of NN. Detailed part of speech abbreviations refer to Table 1 for part of speech explanation.

Formula (5) represents the weights of different templates for different relation categories:

$$\mathrm{P_{pos}^{i}} = \tan h\left(\mathrm{W_H}\left(\mathrm{H_i}\right) + \mathrm{b_H}\right) \tag{5}$$

Where $P_{pos}^{i} = \left\{P_1^i, P_2^i, \ldots, P_n^i\right\}, P_{pos}^{i} \in R^{N \times 1}$, the pos subscript represents part of speech. n is the number of template types. P_n^i represents the weight of the i-th relationship obtained by the n-th Prefix-tuning template. $W_H \in R^{n \times Y}$ and $b_H \in R^{n \times Y}$ are learnable parameter matrices. The subscript H indicates the parameter matrix corresponding to $\mathrm{H_i}$.

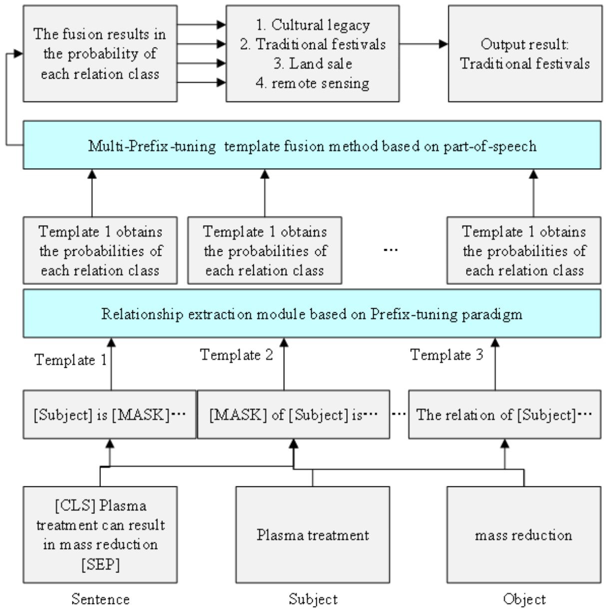

Fig. 1. Multi-Prefix-tuning template fusion method based on part-of-speech

3. Results and discussion

3.1. Data sets and pre-processing

Shu and Farimani [35] has a total of 49 entity relationship categories, which cover a wide range and belong to the general domain of relationship extraction data. This paper randomly extracts some data from the DulE data set for experiments, and randomly divides it into training set, verification set and testing set according to a certain proportion.

Table 3. Parameter setting

Parameter	Value
Word vector dimension	100
Network layer-node number	220
Loss function	Categorical_crossentropy
Learning rate	$1e^{-6}$
Optimizer	Nadam
Dropout	0.5
Epoch	50

COAE-2016-Task3: there are 10 types of entity relation-

Table 4. Performance comparison with different models/%

Data	DuIE			COAE-2016-Task3			FinRE		
Model	P	R	F 1	P	R	F 1	P	R	F 1
NTL	89.95	89.83	89.85	82.51	81.41	80.86	24.12	30.33	20.57
PMRC	93.72	91.73	93.02	76.89	77.13	76.47	36.44	38.66	35.89
TNAPD	93.19	91.73	92.17	86.42	86.03	85.69	38.56	40.67	37.72
MPTTF-BERT	94.38	93.47	93.73	91.71	91.41	91.49	49.46	49.46	49.46

ship in COAE-2016-Task3 dataset [36], which does not belong to a specific domain. Due to the small amount of data in this dataset, 483 unlabeled relationship instances are manually labeled and used as the test set of the experiment.

FinRE dataset [37] is a dataset constructed by manual annotation for completing the English financial relationship extraction task, which contains 44 relationship categories. In this paper, the training set, verification set and testing set used in reference [38] are adopted for experiments.

In this paper, each relationship instance of the above three English relational extraction datasets is preprocessed into the format of {subject, object, relation type, relation text}, and each data set is divided into training set, verification set and testing set during the experiment. The experimental information of the data set is shown in Table 2.

3.2. Experimental setting and evaluation index

The parameter setting of the experiment mainly consists of two parts: model hyperparameter and training parameter. The hyperparameters of the model are mainly the word vector dimension in the input model, the number of nodes in the BiGRU network layer, and the length of the sequence. The length of the sequence is set according to different experimental conditions. During model training, Dropout [38] and early stop training are used to prevent over-fitting of the model. If the loss does not decrease for ten consecutive verification times, the training will be stopped. At the same time, considering the different sizes of data sets, the batch size of COAE-2016-Task3 data set is set to 8 , and the batch size of other data sets is set to 128. Details of parameter settings are shown in Table 3.

In this paper, the effectiveness of the model is measured according to the evaluation index results obtained from the test set. The precision rate (P), recall rate (R) and F1 values are used to quantitatively analyze the model for the classification task of unbalanced data sets. Its calculation is shown in equations (6) $\sim$ (8).

$$\mathrm{P} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}. \tag{6}$$

$$\mathrm{R} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{7}$$

$$\mathrm{F1} = \frac{2\mathrm{PR}}{\mathrm{P}+\mathrm{R}} \tag{8}$$

3.3. Experimental results and comparative analysis

In order to verify the effect of the MPTTF-BERT model proposed in this paper, experiments are conducted to compare it with NTL [39], PMRC [40] and TNAPD [41]. The sequence length of this experiment is set according to the maximum length value of each data set, that is, the sequence length of DuIE, COAE-2016-Task3 and FinRE data sets are set to 160,129 and 96 respectively when making fixed length selection, so as to avoid the situation of reducing the effect of relationship extraction due to the loss of key information. Table 4 shows the relationship extraction effects of the above four experimental models on three English text datasets respectively. The F1 value depends on the results of P and R, and when F1 value is large, it indicates that the experimental method is ideal. It can be seen from Table 6 that the MPTTF-BERT model has the best extraction effect on the three data sets. For the DuIE dataset, the F1 value of MPTTF-BERT reaches 93.73%, and the difference between the F1 value of other models and MPTTF-BERT is as high as 3.88% and as low as 0.71%. On the COAE-2016-Task3 dataset, F1 values of other models differ by 15.02% at the highest and 5.8% at the lowest, indicating that the size of the dataset also affects the learning ability of the model. The larger data amount denotes the stronger learning ability of each model, and the smaller difference in the extraction effect between them. On the FinRE dataset, the F1 value of the MPTTF-BERT model in this paper reaches 49.46%, which is 0.09% higher than that of TNAPD.

The confusion matrix calculated on the test set can reflect the extraction effect of the model in each relation category. The confusion matrix of the MPTTF-BERT model on the COAE-2016-Task3 dataset is shown in table 5, and the confusion matrix on the FinRE dataset is shown in table 6. The integer in the confusion matrix represents the number

Table 5. Confusion matrix of the MPTTF-BERT model on the COAE-2016-Task3

True label	cr2	100	0	0	0	0	0	0	0	0	0
	cr4	0	10	0	1	0	0	0	0	0	0
	cr16	0	0	52	0	0	2	0	0	0	0
	cr20	0	1	0	36	8	2	0	3	0	1
	cr21	0	0	0	2	87	1	0	1	0	0
	cr28	0	0	0	3	3	90	0	4	0	0
	cr29	0	0	0	0	0	0	18	0	0	0
	cr34	0	1	0	4	2	2	0	41	0	0
	cr35	0	0	0	0	0	0	0	0	0	0
	cr37	0	0	0	0	0	1	0	0	0	7
		cr2	cr4	cr16	cr20	cr21	cr28	cr29	cr34	cr35	cr37
		Prediction label									

of instances in which the relationship between the real label and the predicted label exists, and the color depth is displayed. The percentage on the diagonal represents the F1 value calculated by the model after predicting the label.

Table 5 shows that the F1 value corresponding to the real label cr35 is 0.00%, and the number of relationship instances is 0 through data statistics. In addition to the F 1 value of the real label cr20 reaching 74.23%, the F 1 value of other labels is above 82.8%, indicating that the model has a good extraction ability for all relation categories except cr20. By observing Table 6, F1 values of the real tags Unknown, Located, Family and Part-Whole are all relatively high. Statistically, it is found that the number of training instances of the above four tags is relatively large, indicating that the more training instances there are, the better the extraction effect of this model will be.

In the face of the relationship extraction task of unbalanced data sets, it is more accurate to use PR curve to evaluate the model. By calculating the area surrounded by PR curve and X axis, that is, AP value, it can reflect the overall relationship extraction effect of the model on the data set. The larger the AP value is, the better the classification performance of the data set is, and the worse the classification performance is. The PR curve presented by MPTTF-BERT model on these three English text data sets is shown in Figure 2.

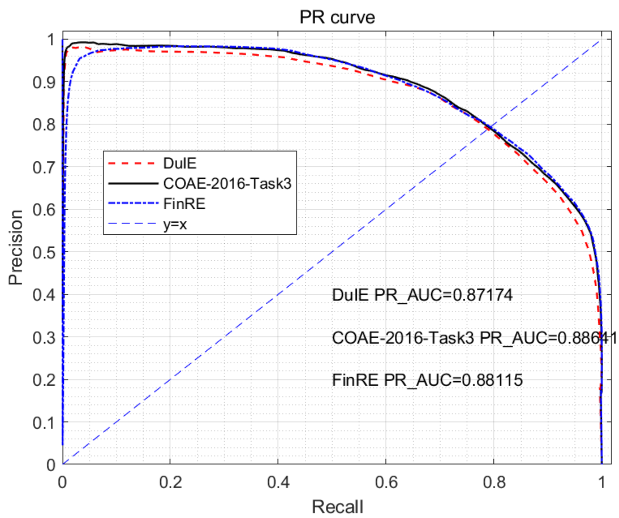

Fig. 2. PR curve on three data sets with MPTTF-BERT model

As can be seen from Figure 2, the MPTTF-BERT model has the best classification effect on the COAE-2016-Task3 dataset with an AUC value of 0.8864 , followed by the FinRE dataset with an AUC value of 0.8811 and the DuIE dataset with a result of 0.8717 . By comparing the above three experimental data sets, it is found that the data sets with better classification effect tend to have less overlap between subject and object, while the DuIE data set has a large number of relation texts with overlap between subject and object, so the classification effect is relatively poor.

In order to illustrate the effectiveness of the Multi-prefix-tuning template and BERT in extracting the relationship features of the model, an ablation experiment is designed to verify it. In this experiment, when preprocessing DuIE and COAE-2016-Task3 data sets, the pre-defined length of their sequences is set to 50 . That is, in the DuIE data set, the number of instances with text length within the pre-defined length accounts for 52.0% of the total instances. Similarly, in the COAE-2016-Task3 dataset, the number of instances with text length within the predefined length accounted for 50.8% of the total instances. The experimental results based on DuIE-50 data set are shown in Table 7, and the experimental results based on COAE-2016-Task3-50 data set are shown in Table 8.

As shown in Table 7 and Table 8, by comparing the experimental results of the first row, the second row and the third row respectively, it can be seen that BERT can effectively improve the effect of entity relationship extrac-

Table 6. Confusion matrix on the FinRE

		un-known	creat	use	near	social	located	owner-ship	general	family	part-whole
True label	unknown	352	2	8	6	13	17	13	8	13	35
	creat	3	11	2	0	2	0	4	0	0	1
	use	21	9	53	0	2	0	10	0	1	7
	near	3	0	0	11	0	5	1	0	0	5
	social	25	0	1	2	68	0	8	6	19	5
	located	89	2	2	30	0	496	3	12	0	70
	owner-ship	8	2	9	0	1	1	23	2	0	2
	general	66	2	5	0	6	1	4	39	2	40
	family	52	0	1	0	12	0	3	3	143	8
	part-whole	41	1	14	2	0	31	12	13	2	213
		Prediction label									

Table 7. Experiments on DulE-50 data set

Model	P/%	R/%	F1/%
Normal BERT	91.26	89.41	90.02
Normal Multi-prefix-tuning template	91.16	89.38	89.94
MPTTF-BERT	93.60	92.34	92.74

Table 8. Experiments on COAE-2016-Task3-50 data set

Model	P/%	R/%	F1/%
Normal BERT	73.60	74.02	73.45
Normal Multi-prefix-tuning template	84.58	84.17	84.15
MPTTF-BERT	87.36	87.48	87.34

tion. The F1 values in the DulE dataset with BERT and the COAE-2016-Task3 are improved by at least 2.67%, and the F1 value in the COaE-2016-Task3 dataset is improved by at least 1.77%. Using the two strategies at the same time can effectively improve the relationship classification ability of the model.

4. conclusions

In the field of RE, there is a problem that it is impossible to label enough training data for all relationships, so zero-shot RE has great research value in this task. The existing RE algorithms based on the Prefix-tuning paradigm cannot be well applied in the zero-shot RE task because of the difficulty in constructing the answer space mapping problem and the need for data resources in the automatic template construction. This paper presents a zero-shot RE model with Multi-prefix-tuning template and BERT. The new model solves the above problems by aligning class representation space with word vector space and using parts of speech to fuse multi-prefix-tuning templates, and introduces the multi-prefix-tuning paradigm into the zero-shot RE task. Finally, several experiments are carried out on the public data set to verify the excellent performance of the proposed model. At present, prefix-tuning is still a discrete word in the zero-shot RE task. In the future, we will further study how to automatically build a continuous prefix-tuning that can efficiently activate the pre-trained language model to further improve the performance of zero-shot RE.

References

[1] Y. Lin, H. Ji, F. Huang, and L. Wu. "A Joint Neural Model for Information Extraction with Global Features." In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 2020, 7999–8009. DOI: 10.18653/v1/2020.acl-main.713.

[2] J. Guo, Y. Fan, L. Pang, L. Yang, Q. Ai, H. Zamani, C. Wu, W. B. Croft, and X. Cheng, (2020) *"A Deep Look into neural ranking models for information retrieval"* **Information Processing & Management 57**(6): DOI: 10.1016/j.ipm.2019.102067.

[3] Z. Yue, L. Hang, Y. Shoulin, and S. Yang, (2018) *"A New Chinese Word Segmentation Method Based on Maximum Matching"* **Journal of Information Hiding and Multimedia Signal Processing.** *9*(6): 1528–1535. DOI: none.

[4] M. Ji, (2021) *"Design of Student Data Acquisition and Extraction Platform Based on Big Data Technology"* **2021 2nd International Conference on Education, Knowledge and Information Management (ICEKIM)**: DOI: 10.1109/ICEKIM52309.2021.00096.

[5] G. Zhiqiang, C. GuoFei, H. Yongming, L. Gang, and L. Fang, (2020) *"Semantic relation extraction using sequential and tree-structured LSTM with attention"* **In-**

formation Sciences *509*: 183–192. DOI: 10.1016/J.INS.2019.09.006.

[6] C. Qu, L. Yang, C. Chen, M. Qiu, W. B. Croft, and M. Iyyer. "Open-Retrieval Conversational Question Answering". In: *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, SIGIR 2020, Virtual Event, China, July 25-30, 2020*. ACM, 2020, 539–548. DOI: 10.1145/3397271.3401110.

[7] H. Guimei and C. Huang. "Deep Information Retrieval-Oriented Smart System for Construction Project Quota Standards". In: *2024 International Conference on Expert Clouds and Applications (ICOECA)*. IEEE, 2024. DOI: 10.1109/ICOECA62351.2024.00023.

[8] L. Xiaofeng, L. Bin, Z. Gang, R. Yibin, Z. Shuangshang, L. Yingjie, G. Le, L. Yuhai, Z. Bin, and W. Fan, (2020) *"Deep learning-based information mining from ocean remote sensing imagery"* **National Science Review 7**: 1584–1605. DOI: 10.1093/nsr/nwaa047.

[9] Z. Lu and Y. Jing, (2021) *"A novel deep CNN method based on aesthetic rule for user preferential images recommendation"* **Journal of Applied Science and Engineering *24***(1): 49–55. DOI: 10.6180/jase.202102_24(1).0006.

[10] H. Yu, H. Li, D. Mao, and Q. Cai, (2020) *"A relationship extraction method for domain knowledge graph construction"* **World Wide Web *23***(2): 735–753. DOI: 10.1007/S11280-019-00765-Y.

[11] Z. Jie, J. Li, and W. Lu. "Learning to Reason Deductively: Math Word Problem Solving as Complex Relation Extraction". In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2022, Dublin, Ireland, May 22-27, 2022*. Association for Computational Linguistics, 2022, 5944–5955. DOI: 10.18653/V1/2022.ACL-LONG.410.

[12] Y. Qin, W. Yang, K. Wang, R. Huang, F. Tian, S. Ao, and Y. Chen, (2021) *"Entity Relation Extraction Based on Entity Indicators"* **Symmetry *13***(4): 539. DOI: 10.3390/SYM13040539.

[13] Z. Geng, Y. Zhang, and Y. Han, (2021) *"Joint entity and relation extraction model based on rich semantics"* **Neurocomputing *429***: 132–140. DOI: 10.1016/J.NEUCOM.2020.12.037.

[14] G. Muhammad, S. Naveed, L. Nadeem, T. Mahmood, A. R. Khan, Y. Amin, and S. A. O. Bahaj, (2023) *"Enhancing Prognosis Accuracy for Ischemic Cardiovascular Disease Using K Nearest Neighbor Algorithm: A Robust Approach"* **IEEE Access *11***: 97879–97895. DOI: 10.1109/ACCESS.2023.3312046.

[15] D. Wang, W. Hu, E. Cao, and W. Sun. "Global-to-Local Neural Networks for Document-Level Relation Extraction". In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020*. Ed. by B. Webber, T. Cohn, Y. He, and Y. Liu. Association for Computational Linguistics, 2020, 3711–3721. DOI: 10.18653/V1/2020.EMNLP-MAIN.303.

[16] M. Miwa and M. Bansal. "End-to-End Relation Extraction using LSTMs on Sequences and Tree Structures". In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers*. The Association for Computer Linguistics, 2016. DOI: 10.18653/V1/P16-1105.

[17] Z. Geng, G. Chen, Y. Han, G. Lu, and F. Li, (2020) *"Semantic relation extraction using sequential and tree-structured LSTM with attention"* **Inf. Sci. *509***: 183–192. DOI: 10.1016/J.INS.2019.09.006.

[18] T. Chen and Y. Hu, (2021) *"Entity relation extraction from electronic medical records based on improved annotation rules and BiLSTM-CRF."* **Annals of Translational Medicine *9***(18): DOI: 10.21037/atm-21-3828.

[19] X. Xu, T. Gao, Y. Wang, and X. Xuan, (2022) *"Event Temporal Relation Extraction with Attention Mechanism and Graph Neural Network"* **Tsinghua Science and Technology *27***(1): 79–90. DOI: 10.26599/TST.2020.9010063.

[20] J. Hou, X. Li, H. Yao, H. Sun, T. Mai, and R. Zhu, (2020) *"BERT-Based Chinese Relation Extraction for Public Security"* **IEEE Access *8***: 132367–132375. DOI: 10.1109/ACCESS.2020.3002863.

[21] T. Nayak and H. T. Ng. "Effective Modeling of Encoder-Decoder Architecture for Joint Entity and Relation Extraction". In: *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*. AAAI Press, 2020, 8528–8535. DOI: 10.1609/AAAI.V34I05.6374.

[22] W. U. Ahmad, N. Peng, and K. Chang. “GATE: Graph Attention Transformer Encoder for Cross-lingual Relation and Event Extraction”. In: *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*. AAAI Press, 2021, 12462–12470. DOI: 10.1609/AAAI.V35I14.17478.

[23] P. H. Cabot and R. Navigli. “REBEL: Relation Extraction By End-to-end Language generation”. In: *Findings of the Association for Computational Linguistics: EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 16-20 November, 2021*. Ed. by M. Moens, X. Huang, L. Specia, and S. W. Yih. Association for Computational Linguistics, 2021, 2370–2381. DOI: 10.18653/V1/2021.FINDINGS-EMNLP.204.

[24] B. Qiao, Z. Zou, Y. Huang, K. Fang, X. Zhu, and Y. Chen, (2022) *“A joint model for entity and relation extraction based on BERT”* **Neural Comput. Appl.** ***34***(5): 3471–3481. DOI: 10.1007/S00521-021-05815-Z.

[25] X. Han and L. Wang, (2020) *“A Novel Document-Level Relation Extraction Method Based on BERT and Entity Information”* **IEEE Access** ***8***: 96912–96919. DOI: 10.1109/ACCESS.2020.2996642.

[26] D. Christou and G. Tsoumakas, (2021) *“Improving Distantly-Supervised Relation Extraction Through BERT-Based Label and Instance Embeddings”* **IEEE Access** ***9***: 62574–62582. DOI: 10.1109/ACCESS.2021.3073428.

[27] C. Chen and C. Li. “ZS-BERT: Towards Zero-Shot Relation Extraction with Attribute Representation Learning”. In: *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2021, Online, June 6-11, 2021*. Association for Computational Linguistics, 2021, 3470–3479. DOI: 10.18653/V1/2021.NAACL-MAIN.272.

[28] J. H. Caufield, H. Hegde, V. Emonet, N. L. Harris, M. P. Joachimiak, N. Matentzoglu, H. Kim, S. Moxon, J. T. Reese, M. A. Haendel, P. N. Robinson, and C. J. Mungall, (2024) *“Structured Prompt Interrogation and Recursive Extraction of Semantics (SPIRES): a method for populating knowledge bases using zero-shot learning”* **Bioinformatics** ***40***(3): btae104. DOI: 10.1093/bioinformatics/btae104.

[29] F. Liu, H. Lin, X. Han, B. Cao, and L. Sun, (2022) *“Pre-training to match for unified low-shot relation extraction”* **arXiv preprint arXiv:2203.12274**: DOI: 10.18653/V1/2022.ACL-LONG.397.

[30] Y. Xiao, C. Tan, Z. Fan, Q. Xu, and W. Zhu. “Joint Entity and Relation Extraction with a Hybrid Transformer and Reinforcement Learning Based Model”. In: *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*. AAAI Press, 2020, 9314–9321. DOI: 10.1609/AAAI.V34I05.6471.

[31] H. Chen, P. Hong, W. Han, N. Majumder, and S. Poria, (2023) *“Dialogue Relation Extraction with Document-Level Heterogeneous Graph Attention Networks”* **Cogn. Comput.** ***15***(2): 793–802. DOI: 10.1007/S12559-023-10110-1.

[32] Y. Zhao, H. Li, and S. Yin, (2022) *“A Multi-channel Character Relationship Classification Model Based on Attention Mechanism”* **International Journal of Mathematical Sciences and Computing(IJMSC)** ***8***(1): 9. DOI: 10.5815/ijmsc.2022.01.03.

[33] K. Zhao, H. Xu, Y. Cheng, X. Li, and K. Gao, (2021) *“Representation iterative fusion based on heterogeneous graph neural network for joint entity and relation extraction”* **Knowl. Based Syst.** ***219***: 106888. DOI: 10.1016/J.KNOSYS.2021.106888.

[34] Q. Li, Y. Wang, T. You, and Y. Lu, (2022) *“BioKnowPrompt: Incorporating imprecise knowledge into prompt-tuning verbalizer with biomedical text for relation extraction”* **Inf. Sci.** ***617***: 346–358. DOI: 10.1016/J.INS.2022.10.063.

[35] D. Shu and A. B. Farimani, (2024) *“Zero-Shot Uncertainty Quantification using Diffusion Probabilistic Models”* **CoRR** ***abs/2408.04718***: DOI: 10.48550/ARXIV.2408.04718. arXiv: 2408.04718.

[36] D. Peng, D. Zhang, C. Liu, and J. Lu, (2020) *“BG-SAC: Entity relationship classification model based on Self-Attention supported Capsule Networks”* **Appl. Soft Comput.** ***91***: 106186. DOI: 10.1016/J.ASOC.2020.106186.

[37] Y. Zhang, (2023) *“Relation extraction in Chinese using attention-based bidirectional long short-term memory networks”* **PeerJ Computer Science** ***9***: e1509.

[38] S. Yin, H. Li, A. A. Laghari, T. R. Gadekallu, G. A. R. Sampedro, and A. S. Almadhor, (2024) *"An Anomaly Detection Model Based on Deep Auto-Encoder and Capsule Graph Convolution via Sparrow Search Algorithm in 6G Internet of Everything"* **IEEE Internet Things J.** ***11***(18): 29402–29411. DOI: 10.1109/JIOT.2024.3353337.

[39] H. Fabregat, A. Duque, J. Martínez-Romo, and L. Araujo, (2023) *"Negation-based transfer learning for improving biomedical Named Entity Recognition and Relation Extraction"* **J. Biomed. Informatics** ***138***: 104279. DOI: 10.1016/J.JBI.2022.104279.

[40] C. Peng, X. Yang, Z. Yu, J. Bian, W. R. Hogan, and Y. Wu, (2023) *"Clinical concept and relation extraction using prompt-based machine reading comprehension"* **J. Am. Medical Informatics Assoc.** ***30***(9): 1486–1493. DOI: 10.1093/JAMIA/OCAD107.

[41] D. Sui, X. Zeng, Y. Chen, K. Liu, and J. Zhao, (2024) *"Joint Entity and Relation Extraction With Set Prediction Networks"* **IEEE Trans. Neural Networks Learn. Syst.** ***35***(9): 12784–12795. DOI: 10.1109/TNNLS.2023.3264735.