

BERT-PAGG: a Chinese relationship extraction model fusing PAGG and entity location information

Bin Xu, Shuai Li, Zhaowu Zhang and Tongxin Liao

School of Computer Science and Engineering, Northeastern University, Shenyang, Liaoning, China

ABSTRACT

Relationship extraction is one of the important tasks of constructing knowledge graph. In recent years, many scholars have introduced external information other than entities into relationship extraction models, which perform better than traditional relationship extraction methods. However, they ignore the importance of the relative position between entities. Considering the relative position between entity pairs and the influence of sentence level information on the performance of relationship extraction model, this article proposes a BERT-PAGG relationship extraction model. The model introduces the location information of entities, and combines the local features extracted by PAGG module with the entity vector representation output by BERT. Specifically, BERT-PAGG integrates entity location information into local features through segmented convolution neural network, uses attention mechanism to capture more effective semantic features, and finally regulates the transmission of information flow through gating mechanism. Experimental results on two open Chinese relation extraction datasets show that the proposed method achieves the best results compared with other models. At the same time, ablation experiments show that PAGG module can effectively use external information, and the introduction of this module makes the Macro-F1 value of the model increase by at least 2.82%.

Subjects Artificial Intelligence, Data Mining and Machine Learning, Text Mining

Keywords Relation extraction, Entity type, Segmented convolutional neural networks, Gating mechanism, Self-attention mechanism

INTRODUCTION

Relationship extraction aims to find the semantic relationship between two related entities, which may be unrelated or related in some way. In this article, we focus on Chinese relationship extraction. For Chinese relational extraction, the linguistic and semantic environment is more complex and it typically faces two challenges: (1) wrong participle. The example sentence “化妆和服装” has the correct split result of “化妆/和/服装”, but if it is split into “化妆/和服/装”, the semantic meaning is altered, which will affect the subsequent parsing work. (2) Chinese has a more flexible grammatical structure and a more variable sentence structure. As a result, more grammatical structures and more meaningful features must be considered in Chinese relational extraction.

The traditional approach to the relation extraction task is based on deep learning to obtain information about entities in sentences *via* neural networks and subsequently

Submitted 23 March 2023

Accepted 8 June 2023

Published 17 July 2023

Corresponding author

Bin Xu, xubin@mail.neu.edu.cn

Academic editor

Xiangjie Kong

Additional Information and
Declarations can be found on
page 13

DOI 10.7717/peerj-cs.1470

© Copyright
2023 Xu et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

perform relation extraction, but this approach does not take into account other information contained in sentences, such as sentence information, entity location information, and syntactic structure information. Another key factor in improving the relationship extraction effect is the use of this external information. For example, the Chinese sentences “奥巴马出生在美国” and “奥巴马在美国曾担任总统职位”, both sentences have the entities “奥巴马” and “美国”, but the relationship between the entities in the two sentences is completely different, and the relationship between the different sentences containing the same entities may be different or the same. If we simply use entity information for relationship extraction and use the same word vector for entities, it will cause difficulties in relationship classification and affect the effectiveness of relationship extraction. Therefore, how to represent entities and introduce sentence information in the relationship extraction model are the key factors to improve the performance, and good entity representation and rich sentence information can avoid the above problems.

Furthermore, it can be found from the above two sentences that the relationship between entities in Chinese sentences is mostly the content between entities, and the positional relationship between characters other than entities and entities is also a key factor affecting the performance of relationship extraction. For example, the word “出生” in the above sentence is located between two entities, and it is essential to introduce such location information. Note that the syntactic structure should also be taken into account in the relationship extraction task, since the features extracted using word sequences alone are not sufficient and cannot accurately model the relationships between entities, and a model using syntactic dependencies can capture more useful information.

In this article, we propose the BERT-PAGG (Bert-based Piecewise Convolutional Neural Networks, Self-Attention Gate and GCN model) relationship extraction model, which improves the relationship extraction by using a combination of external and entity information in order to extraction performance by using a combination of external and entity information. The main work is as follows.

1. We add “#” and “\$” signs on both sides of the head and tail entities of the input text to make the BERT pre-trained model better capture inter-entity information.
2. To better capture the structural information features of the two entities, we introduce the relative positions of the entities and extract the local features of the sentences by a segmented convolutional neural network PCNN.
3. The self-attention mechanism is employed to capture global dependent features and a gating mechanism is utilized to adaptively fuse global and local features.
4. A graph convolutional neural network is employed to fuse semantic and structural information with entity-specific masks to mask the non-entity words learned by the GCN layer, followed by the use of a retrieval-based attention mechanism to extract important features of contextual words, which are finally combined with the features output by the gating mechanism for the relationship extraction task.

RELATED WORK

As an essential subtask in information extraction, relation extraction is also a key step in building knowledge graphs. With the emergence of deep learning, existing relationship

extraction models mainly rely on neural networks to extract the semantic information of sentences.

Socher et al. (2012) combined a matrix vector representation with a recurrent neural network. The model learns a combined vector representation of phrases and sentences of arbitrary syntactic type and length. The problem that the word vector space model cannot capture long sentence information is solved. However, since RNN cannot process the long-term dependency problem and the existence of gradient explosion problem, *Le et al. (2018)* innovatively proposed a multichannel long short-term memory (LSTM) model, which combines multiple feature embeddings and uses BiLSTM to encode different feature embeddings to improve the adaptability and robustness.

CNN can extract local features more efficiently in shorter sentences and has a simpler structure and faster training than RNN. *Zeng et al. (2014)* used convolutional neural networks for the first time to extract hierarchical features of words and sentences for relationship extraction in a relationship classification task and used positional features to improve feature extraction, resulting in a significant improvement in model performance. *Nguyen & Grishman (2015)* used convolutional neural networks to replace the traditional feature engineering approach, using convolutional kernels of different sizes to extract local features of entity contexts and using positional embeddings to encode relative distances, resulting in a further improvement in the performance of the proposed model over the single kernel CNN. To reduce the impact of error accumulation kernel propagation, *Zeng et al. (2015)* improved the pooling layer of traditional convolutional neural network (CNN) to obtain PCNN (Piecewise CNN), which pools the feature map into three segments by two entity locations, with the aim of better capturing the structured information between two entities and using attention mechanism by building sentence level selective attention neural model to alleviate the mislabeling problem, these efforts improve the overall performance of the model.

As noisy signals will cause errors, in order to reduce their impact on model performance, *Lin et al. (2016)* applied the attention mechanism to PCNN. They used a convolutional neural network to embed the semantics of sentences and constructed sentence-level attention on instances to reduce the impact of noise by adjusting the weights. Although the PCNN with the introduction of the attention mechanism has achieved some success in the relational extraction task, it does not have good access to contextual information; therefore, *Lee, gyu Seo & Choi (2019)* proposed a new end-to-end recurrent neural model that combines LSTM and entity-aware attention mechanism, which enables the model to utilize contextual information and solve the long-term dependency problem with better results. While *Zhou & Chen (2022)* improved entity representation to reduce noise in labels and better reveal the contextual semantic representation of entities.

In order to make the model utilize more information, many scholars started to incorporate features of different granularity into the model, such as word-level features and character-level features. *Zhang & Yang (2018)* proposed a lattice-LSTM model, which encodes a series of input characters as well as all potential words from the matching dictionary, where the gated recurrent cell both makes full use of the word sequence information and enables the model to select the most relevant words and characters

from the sentence. *Xu, Yuan & Zhong (2020)* merged word and character inputs and proposed a Lattice-GRU model with Bi-GRU instead of Bi-LSTM. *Li et al. (2019)* proposed a model for Chinese relation extraction that incorporates word-level information into character sequences and introduces external knowledge. *Huang et al. (2021)* used the intermediate layer of BERT to obtain different levels of semantic information and design multi-granularity features for the final relation classification. *Kong et al. (2021)* proposed an adaptive approach to improve the performance of Chinese relationship extraction by combining word information into the embedding layer of the character input-based model using a dictionary to address the information loss problem of the MG-Lattice model. *Xue et al. (2019)* proposed a focused attention model based on the BERT language model and dynamic range attention mechanism for joint entity and relation extraction tasks. The model combines the BERT language model with joint learning to improve the feature representation capability of the shared parameter layer through a dynamic range attention mechanism.

In recent years, some researchers have considered the semantic connection between entity positions. *Zeng et al. (2014)* introduced positional embedding, which takes the relative distance between each word and the target entity as an important input. Considering the importance of syntactic dependencies, researchers have improved LSTM, and *Tai, Socher & Manning (2015)* proposed TreeLSTM, through which the tree structure model is able to use the syntactic structure tree of a sentence as an input to obtain sentence features for relation extraction. The graph neural network proposed by *Scarselli et al. (2009)* has received wide attention from researchers, and *Zhang, Qi & Manning (2018)* were the first to use graph convolutional networks for the relational extraction task, constructing graphs through the dependency tree structure of sentences and performing path-centered pruning on the dependency trees, and proposing the C-GCN model based on the PA-LSTM (*Zhang et al., 2017*) model for the relational extraction task significant progress was achieved. Subsequently, based on the study of C-GCN by *Guo, Zhang & Lu (2019)* and others, the AGGCN model was proposed, where instead of pruning the dependency tree afterward as the input to the model, the model was dynamically adjusted by the Attention mechanism, by which the model was able to capture richer local and non-local information.

Since Chinese texts from different domains involve widely different relationship and entity features, and the high complexity of the Chinese language, including the complexity of language structure, semantic complexity, and vocabulary complexity, all these bring challenges to Chinese relationship extraction. Furthermore, since texts oriented to the social media domain often lack valid information, finding content beyond the text to supplement textual information is imperative, and according to the review by *Zhu et al. (2023)* and *Zhu et al. (2022)*. multimodal information extraction is an important research area that many scholars have now applied to specific domains.

BERT-PAGG MODEL

To perform better relationship extraction, the model proposed in this article introduces a segmental convolutional neural network, self-attention mechanism, gating mechanism, and

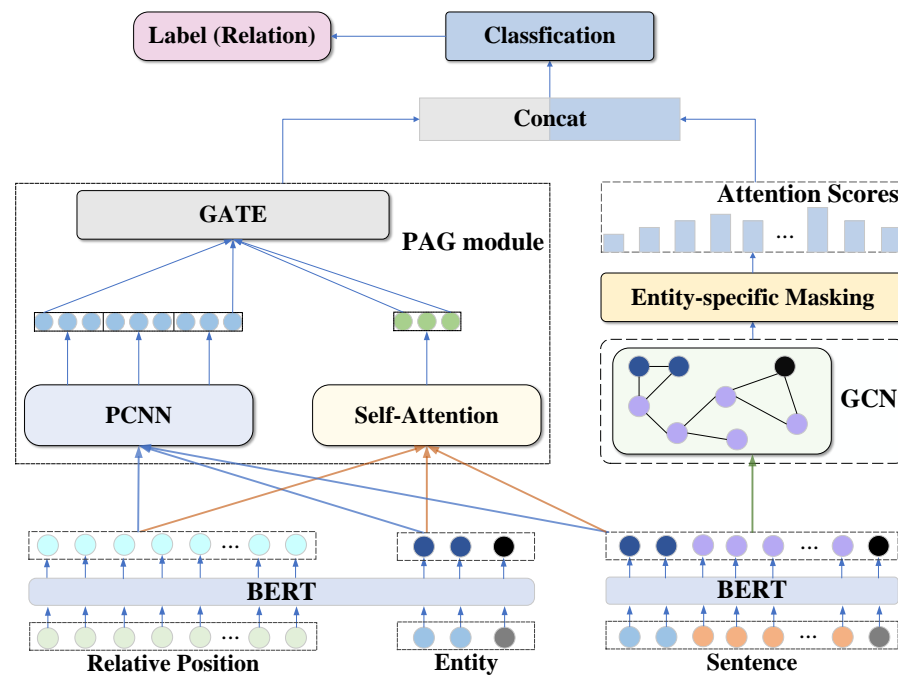


Figure 1 Overview of BERT-PAGG structure.

Full-size DOI: [10.7717/peerjcs.1470/fig-1](https://doi.org/10.7717/peerjcs.1470/fig-1)

graph neural network based on BERT pre-training model, and implements a BERT-PAGG relationship extraction model based on BERT. The overall structure of the model is shown in Fig. 1.

Input layer

The input of the BERT-PAGG model mainly consists of sentences $S = \{c_1, c_2, \dots, c_n\}$ and entity positions. Firstly, the input sequence $\{c_1, c_2, \dots, c_n\}$ is encoded by the BERT pre-training model to obtain $\{e^{c_1}, e^{c_2}, \dots, e^{c_n}\}$, where each word c_i in the sentence corresponds to the encoded e^{c_i} .

Vector representation layer

The model proposed in this article is also built on the BERT pre-training model, which can obtain richer information from the text and provide better text representation. In the relational extraction task, how to obtain entity information efficiently is the key to improving the performance of the relational extraction model. *Wu & He (2019)* proposed the R-BERT relational extraction model for the relational classification task, and proved that the tagging process for entities in the relational extraction model is effective. Special tags, such as “#” and “\$”, are added to the head and tail entities in the input text, and such tags enable the BERT pre-training model to capture entity information better.

For the text in the dataset, special markers are added to the head and tail entities according to their positions, respectively, in the form shown in Table 1.

Wu & He (2019) conducted ablation experiments for the method of adding markers to the head and tail entities, and the experimental results showed that the effect of adding

Table 1 Example of adding special tags into text.

Dataset	Sentence	Relation
FinRE	# 东方航空 #AH 股临时停牌传将与 \$ 上航 \$ 合并。 Eastern Airlines # Stock suspension \$ Shanghai Airlines \$ merge.	合并 merge
SanWen	# 旅行车 # 继续在 \$ 崇山峻岭 \$ 中穿行。 #station wagon # continue \$ mountains\$ travel through.	Located

markers to the head and tail entities separately is better than that of adding markers to the head or tail entities, and also better than the effect of not adding markers. Therefore, in this article, we add “#” and “\$” marks on both sides of the head and tail entities of the input text, which enables the Attention mechanism of the Transformer structure in the BERT pre-training model to better capture the information of the two entities and the relationship information between them, including the relationship information between the two entities, the information between the two entities and other words, and the semantic information between the two entities.

We input the above text sequence with “#” and “\$” marks and entity positions into BERT to obtain the corresponding entity vector representation, where the vector representation of the first word of the head entity and the vector representation of the first word of the tail entity are spliced to represent the vector representation of the head and tail entities as shown in Eq. (1).

$$Entity = [h_{head}; h_{end}]. \quad (1)$$

Segmented convolutional neural network

To highlight the importance of the first and last entities and their location information, we use a segmental convolutional neural network (PCNN) to extract sentence local features. Segmented convolutional neural network PCNN is a widely used model for relation extraction. Considering that the relative positions of words and entities also have an impact on the effect of relation extraction in the relation extraction task, the position information is also added to the encoding layer. Since the traditional single max-pooling can only get one value per convolutional kernel, which cannot capture the structural information features between two entities, PCNN introduces a segmented max-pooling mechanism, which cuts the text into three parts according to the head entity and the tail entity, extracts the local features separately and then splices the three vectors obtained by segmented max-pooling to obtain the final feature vector, This mechanism can extract the sentence feature information more effectively. The structure of the PCNN network is shown in Fig. 2.

The PCNN relationship extraction model differs from the traditional relationship extraction model by using only the word vector representation as the input of the model. Considering the importance of the positional relationship between words and entities, PCNN distinguishes entities from other words and also splices the positional features of words and head and tail entities with word embedding in the vector representation layer

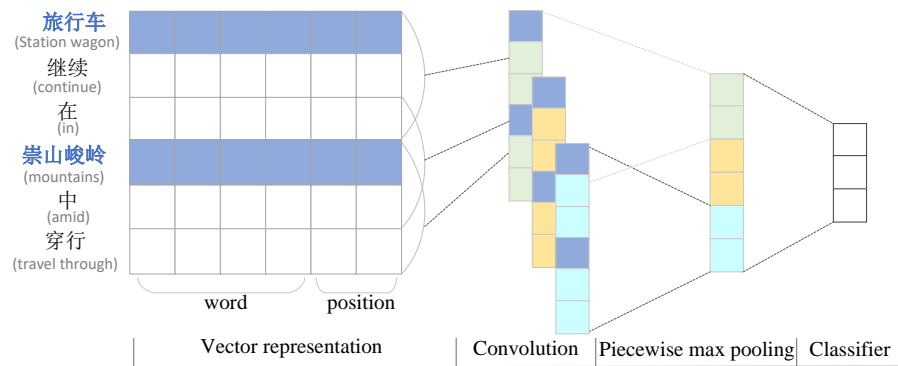


Figure 2 Diagram of piecewise convolutional neural network.

Full-size DOI: [10.7717/peerjcs.1470/fig-2](https://doi.org/10.7717/peerjcs.1470/fig-2)

as the input of the final model, enabling the model to obtain rich positional information. The calculation is shown in Eqs. (2)–(4).

$$B = BERT(S) \quad (2)$$

$$Position = Embedding(Index) \quad (3)$$

$$P = PCNN(B, Position) \quad (4)$$

where B is the embedding representation of the sentence S after BERT encoding, $index$ is the position of the word relative to the head and tail entities, the corresponding position embedding representation is obtained by *Embedding*, and finally the word embedding B and position embedding $Position$ are stitched together and input to PCNN to get the output P .

Through the studies and analysis of the distance between the word and the entity in the sentence, it was found that the smaller the distance between the word and the entity, the greater the influence on the final relationship extraction results, so the relative distance between the word and the head and tail entity was used to represent the location feature in PCNN. As shown in Table 2, the head entity in the sentence is “旅行车” and the tail entity is “崇山峻岭”, and the location features of other characters in the sentence are expressed as the relative distance from the current word to the head and tail entities, which are encoded by the location representation matrix $D \in R^{d^p \times |v^p|}$, where d^p is the dimension of the position vector, and v^p denotes a fixed-size position table.

The word vector representation and the position vector representation are concatenated to form the final vector representation as the input of the model, for the input text $S = \{c_1, c_2, \dots, c_n\}$, the word vector representation $\{e^{c_1}, e^{c_2}, \dots, e^{c_n}\}$ is obtained by BERT, and the position vector representations are obtained by position features. The final stitching yields $SP = [S; HEAD; TAIL]$, where $SP \in R^d$ ($d \in d^w + 2 \times d^p$) is taken as the input to the model.

$$PCNN_{\text{outputs}} = \tanh(P_{1:n}). \quad (5)$$

Table 2 Relative head and tail entity position.

Sample	旅行车 Station wagon	继 续 Continue	在 In	崇山峻岭 Mountains	中 Amid	穿 Travel	行 Through	
Relative head entity position	0	1	2	3	4	5	6	7
Relative tail entity position	-4	-3	-2	-1	0	1	2	3

Self-attention mechanism

The PCNN has the advantages of capturing local features, small computation and simple structure, and is the mainstream model in relation extraction. The contextual representation of sentences can be obtained by using PCNN, but long-term dependency features of sentences cannot be obtained by using PCNN alone, even by stacking multiple convolutional neural networks to achieve better performance. In this article, the self-attention mechanism is employed to capture global dependency features that complement the contextual representation obtained by PCNN.

Instead of using the PCNN in tandem with the self-attention mechanism in the previous way, a parallel approach is adopted in this article. For the sentence input, the contextual representation of the sentence is obtained by PCNN, the global dependent features are captured by the self-attention mechanism, and finally the global dependent features and the contextual representation are put into the gating mechanism to adaptively fuse the two features and get the final output. The PCNN has been introduced in the previous section, and the output calculation formula of the self-attention mechanism module is given here as Eqs. (6)–(8) shows.

$$Att = W^2 \sigma(W^1 SP + b^1) + b^2 \quad (6)$$

$$P^{(Att)} = \text{softmax}(Att) \quad (7)$$

$$A = \sum P^{(Att)} \odot SP \quad (8)$$

where W^1 and W^2 are the weight matrices, b^1 and b^2 are the biases, and the final output A is obtained.

Gating mechanism

To introduce both local and global dependent features of a sentence into the relational extraction task, we need to fuse the two features. Traditional features fusion approaches (e.g., feature stitching (Bell et al., 2015), feature summation (Kong et al., 2016), etc.) do not consider whether the participating features are useful for the final task, so the fused features may contain a lot of unwanted information. Considering that the gating mechanism based on Gated Tanh Units (GTU) and Gated Linear Units (GLU) will measure the usefulness of each feature vector in the feature map during fusion and select the information that is useful for the final task for aggregation, therefore, this article uses the gating mechanism to adaptively fuse the entity features extracted by the PCNN network and the global dependent features extracted by the attention mechanism are fed to the classification

layer as a representation of the whole text for classification, and the gating mechanism is calculated as shown in Eqs. (9)–10.

$$g = \text{sigmoid}(W^{g1} \tanh(W^{g2}A + b^{g2}) + b^{g1}) \quad (9)$$

$$\text{outputs} = g \cdot \text{PCNN}_{\text{outputs}} \quad (10)$$

where W^{g1} and W^{g2} are the weight matrices, b^{g1} and b^{g2} are the biases, and finally the *outputs* is obtained through the gating mechanism.

Entity-specific graph convolutional neural networks

Constructing the graph

To construct the graph neural network structure in the relational extraction task, we construct a syntactic dependency graph for each sentence. Figure 3 shows the construction process from word dependency to adjacency matrix. Specifically, we use Stanza to obtain the syntactic dependency tree of the sentence, and based on the syntactic dependency tree, we can construct the adjacency matrix A of the sentence. The adjacency matrix A is defined as shown in Eq. (11).

$$A_{i,j} = \begin{cases} 1, & \text{if } c_i, c_j \text{ has dependency} \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

where we set the main diagonal of the adjacency matrix A to 1 in order to preserve the information of its own nodes, and we construct the undirected graph, *i.e.*, $A_{i,j} = A_{j,i}$, since we believe that the directed graph will lose some of the dependency information.

Graph convolutional neural network

We feed the adjacency matrix A constructed above into the graph convolutional network, which in this article is stacked in multiple layers, and then update the node representation in the next layer based on the hidden representation of the node's neighborhood, with the graph convolutional formula shown in Eq. (12).

$$h_i^l = \text{relu} \frac{A_i h_i^{l-1} W^l}{D+1} + B^l \quad (12)$$

$$D_{i,j} = \begin{cases} 1, & A_{i,j} \neq 0 \\ 0, & A_{i,j} = 0 \end{cases} \quad (13)$$

where *relu* is the nonlinear activation function, W^l is the linear transformation weight, B^l is the linear transformation bias term, $D = \sum_{j=1}^n D_{i,j}$ is the degree of A_i , and h_i^{l-1} is the hidden representation of the GCN transformed from the previous layer.

Entity-specific masking mechanism

Above we capture the regional features of neighboring nodes by stacking multiple layers of GCNs, and to retrieve the semantic relationships of entities on their contextual words,

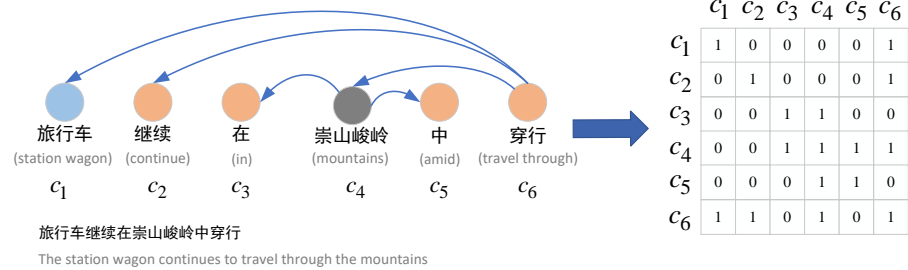


Figure 3 Construction of adjacency matrix.

Full-size DOI: [10.7717/peerjcs.1470/fig-3](https://doi.org/10.7717/peerjcs.1470/fig-3)

we use an entity-specific masking mechanism, specifically, we mask the output non-entity vector of the last layer of GCNs with 0, while the vector representation of entities remains unchanged, as shown in Eq. (14).

$$Mask_i^l = 0 \quad 0 \leq i < h, h+1 \leq i < t, t+1 \leq i < n \quad (14)$$

where h is the position index of the head entity, while t is the position index of the tail entity and n is the sentence length. Thus, we can obtain a representation of the entity-specific mask:

$$h_{mask}^L = \{0, \dots, h_h^L, \dots, h_t^L, \dots, 0\}. \quad (15)$$

Similar to [Hong et al. \(2021\)](#), we use the dot product attention mechanism to extract entity-specific features from contextual information and set the corresponding attention weights for each contextual word, and the attention weights are calculated under.

$$\beta_i = \sum_{j=1}^n \mathbf{h}_{i0} \top h_j^L \quad (16)$$

$$\alpha_i = \frac{\exp(\beta_i)}{\sum_{j=1}^n \exp(\beta_j)}. \quad (17)$$

The dot product can calculate the semantic correlation between entities and contextual words, which leads to the computation of the attention weight matrix, so that the final calculation of the input vector is shown in Eq. (18).

$$R_{gcn} = \sum_{i=1}^n \alpha_i e^{c_i}. \quad (18)$$

Classification layer

Formally, we first fuse the local features extracted by PCNN and the global features focused on by the self-attention mechanism using the gating mechanism, and subsequently splice the fused features with the dependent features extracted by GCN, which in turn yields

the final representation that is fed into the classification layer to produce the probability distribution p , as shown in Eqs. (19)–(20).

$$r = [R_{gate}; R_{gcn}] \quad (19)$$

$$p = fc(\text{relu}(W_p r + b_p)) \quad (20)$$

where ‘;’ denotes the splicing of vectors, fc is the fully connected function, W_p and b_p are trainable parameters.

EXPERIMENT

Datasets and setup

This article evaluates the BERT-PAGG model on two public datasets: the SanWen dataset and the FinRE dataset, in which the SanWen dataset collects a corpus of 837 Chinese prose works, including 695 training sets, 58 validation sets and 84 test sets. The dataset includes 10 types of relations including “Unknown”. While the FinRE dataset collected 2,647 Sina financial news, and extracted 13,486 relationship instances as the training set, 3,727 relationship instances as the test set, and 1,489 relationship instances as the validation set, which contains 44 types of relationships.

In this article, a 768-dimensional word vector is generated using the BERT model with Adam as the optimizer and the learning rate set to $5e-5$. To prevent overfitting, the dropout strategy is applied to the input and hidden layers in the training model and the dropout rate is set to 0.1. The hyperparameter settings of the BERT-PAGG model are shown in Table 3.

Baselines

To evaluate the performance of the BERT-PAGG model for Chinese relationship extraction task, this article compares it with the following classical models.

- **Att-BLSTM:** *Zhou et al. (2016)* proposed a BLSTM model with word-level attention mechanism.
- **BLSTM:** A bidirectional LSTM model for relation extraction proposed by *Zhang & Wang (2015)*.
- **PCNN:** A piecewise convolutional neural network model with multi-instance learning proposed by *Le et al. (2018)*.
- **PCNN+Att:** An improved PCNN model with selective attention mechanism proposed by *Lee, gyu Seo & Choi (2019)*.
- **Lattice LSTM:** A model for Chinese named entity recognition proposed by *Zhang & Yang (2018)*, which explicitly utilizes word and word order information.
- **MG Lattice:** A model for Chinese relation extraction proposed by *Li et al. (2019)*, which incorporates word-level information into character sequences and introduces external knowledge.
- **MGRSA:** *Zhong (2021)* proposed a Chinese entity relationship extraction model based on a multi-level threshold recursion mechanism and self-attention mechanism.

Table 3 Hyper parameters of BERT-PAGG model.

Parameter	Value
batch_size	64
epoch	10
lr	5e-5
optimizer	Adam
dropout	0.1
GCN-Layers	3

Table 4 Overall performance of different models on relationship extraction tasks.

Model	SanWen Macro-F1	FinRE Macro-F1
Att-BLSTM	59.48	41.48
PCNN+Att	60.55	46.13
BLSTM	61.04	42.87
PCNN	61.20	45.51
Lattice LSTM	63.88	47.41
MG Lattice	65.61	49.26
MGRSA	67.12	52.61
ExSoftwords	70.90	—
BERT-PAGG	73.83	53.01

Notes.

The best results are in bold.

Main results

As shown in Table 4, we report the performance of the BERT-PAGG model and the baseline model on two datasets, where the BERT-PAGG model is marked with “#” and “\$” on both sides of the head and tail entities in the sample, respectively.

As can be seen from the above table, our proposed BERT-PAGG shows effectiveness and superiority in the relational extraction task compared to others, achieving a state-of-the-art Macro-F1 score. First, piecewise convolutional neural networks and self-attention mechanism can capture local features and global features well, while the addition of graph neural networks allows inter-word dependency features to be used effectively. Thus, we can observe that BERT-PAGG achieves the highest Macro-F1 value compared to the model using only sequence information.

Ablation study

To investigate the effect of different components, we performed ablation experiments, and the results are shown in Table 5, where rows 1-4 are removing the specified components separately, in row 5, we use the lightweight Bert model, *i.e.*, Albert, to replace Bert to verify the effectiveness of BERT. From the Table 5, we can draw the following conclusions.

1. The attention mechanism and the PCNN module can improve the feature extraction, thus improving the performance of the Chinese relationship extraction model.

Table 5 The ablation study results.

Variants	SanWen Macro-F1(%)	FinRE Macro-F1(%)
M_0 : w/o Self-Attention	72.68	48.80
M_1 : w/o PCNN	72.45	48.45
M_2 : w/o Gate	70.76	48.27
M_3 : w/o GCN	73.26	52.24
M_4 : Only replace Bert with Albert	63.89	39.40
M_5 : BERT-PAGG (full model)	73.83	53.01

Notes.

The best results are in bold.

- The introduction of gating mechanism can better integrate the features extracted by the attention mechanism and PCNN module, and the gating mechanism can filter out the key features more wisely. Removing the gating mechanism reduces the BERT-PAGG model by at least 3% in the Macro-F1 metric, which indicates that the representation of text by PCNN and self-attention can be better obtained by the gating mechanism.
- Without using inter-word dependency information, *i.e.*, removing the GCN module, the Macro-F1 of M_3 is lower than that of M_5 , this shows that the GCN module can effectively extract the dependency information between words and apply it to the relation extraction task. However, it is worth noting that the drop in Macro-F1 score is not significant after removing the GCN module, which may be due to the poor accuracy of the syntactic analyzer in the Chinese context, which in turn leads to errors in the construction of the adjacency matrix.
- To verify the effectiveness of BERT for the Chinese relationship extraction task, M_4 uses Albert with fewer parameters to replace BERT. Compared with M_5 , M_4 has a significant decline in Macro-F1, which indicates that BERT pre-training model can obtain more information from text and provide better text representation.

CONCLUSION

In the experiments of relationship extraction, the PAGG module is designed in this article, which enables the model to incorporate entity location information and local features as well as dependency features, and pay more attention to key information, so that the BERT-PAGG relation extraction model proposed in this article achieves performance improvement compared with traditional relationship extraction models, such as LSTM, MG Lattice, PCNN, MGRSA, Att-BLSTM, etc. Meanwhile, to prove the effectiveness of the PAGG module designed in this article, the ablation experimental results show that the introduction of the PAGG module enhances the performance of relation extraction.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by the Fundamental Research Funds for the Central Universities (N2116019), the Liaoning Natural Science Foundation (2022-MS-119), the National Natural Science Foundation of China (Grant No. 62137001), the National Natural Science Foundation of China, grant number 72271048 and the Liaoning Province Discipline Inspection Supervision Big Data Key Laboratory ,grant number ZX20220460. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

The Fundamental Research Funds for the Central Universities: N2116019.

The Liaoning Natural Science Foundation: 2022-MS-119.

The National Natural Science Foundation of China: 62137001.

The National Natural Science Foundation of China: 72271048.

The Liaoning Province Discipline Inspection Supervision Big Data Key Laboratory: ZX20220460.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Bin Xu conceived and designed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Shuai Li conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Zhaowu Zhang conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Tongxin Liao conceived and designed the experiments, performed the computation work, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The source code and data can be accessed at Zenodo: lishuaics, & Wenxuan Zhou. (2023). L-shuai/BERTPAG_GCN_final: v1.0.0 (dev). Zenodo. <https://doi.org/10.5281/zenodo.7962611>.

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj-cs.1470#supplemental-information>.

REFERENCES

- Bell S, Zitnick CL, Bala K, Girshick RB. 2015.** Inside-outside net: detecting objects in context with skip pooling and recurrent neural networks. In: *2016 IEEE conference on computer vision and pattern recognition (CVPR)*. Piscataway: IEEE, 2874–2883.
- Guo Z, Zhang Y, Lu W. 2019.** Attention guided graph convolutional networks for relation extraction. ArXiv preprint. [arXiv:1906.07510](https://arxiv.org/abs/1906.07510).
- Hong W, Li S, Hu Z, Rasool A, Jiang Q, Weng Y. 2021.** Improving relation extraction by knowledge representation learning. In: *2021 IEEE 33rd international conference on tools with artificial intelligence (ICTAI)*. 1211–1215 [DOI 10.1109/ICTAI52525.2021.00191](https://doi.org/10.1109/ICTAI52525.2021.00191).
- Huang Y, Li Z, Deng W, Wang G, Lin Z. 2021.** D-BERT: incorporating dependency-based attention into BERT for relation extraction. *CAAI Transactions on Intelligence Technology* 6(4):417–425 [DOI 10.1049/cit2.12033](https://doi.org/10.1049/cit2.12033).
- Kong B, Liu S, Wei F, Liruizhi J, Wang G. 2021.** Chinese relation extraction using extend softword. *IEEE Access* PP:1–1 [DOI 10.1109/ACCESS.2021.3102225](https://doi.org/10.1109/ACCESS.2021.3102225).
- Kong T, Yao A, Chen Y, Sun F. 2016.** HyperNet: towards accurate region proposal generation and joint object detection. In: *2016 IEEE conference on computer vision and pattern recognition (CVPR)*. Piscataway: IEEE, 845–853.
- Le H-Q, Can D-C, Vu ST, Dang TH, Pilehvar MT, Collier N. 2018.** Large-scale exploration of neural relation classification architectures. In: *Proceedings of the 2018 conference on empirical methods in natural language processing*. Brussels, Belgium: Association for Computational Linguistics, 2266–2277 [DOI 10.18653/v1/D18-1250](https://doi.org/10.18653/v1/D18-1250).
- Lee J, Gyu Seo S, Choi YS. 2019.** Semantic relation classification via bidirectional LSTM networks with entity-aware attention using latent entity typing. ArXiv preprint. [arXiv:1901.08163](https://arxiv.org/abs/1901.08163).
- Li Z, Ding N, Liu Z, Zheng H, Shen Y. 2019.** Chinese relation extraction with multi-grained information and external linguistic knowledge. In: *Annual meeting of the association for computational linguistics*.
- Lin Y, Shen S, Liu Z, Luan H, Sun M. 2016.** Neural relation extraction with selective attention over instances. In: *Proceedings of the 54th annual meeting of the association for computational linguistics (volume 1: long papers)*. Berlin, Germany: Association for Computational Linguistics, 2124–2133 [DOI 10.18653/v1/P16-1200](https://doi.org/10.18653/v1/P16-1200).
- Nguyen TH, Grishman R. 2015.** Relation extraction: perspective from convolutional neural networks. In: *Proceedings of the 1st workshop on vector space modeling for natural language processing*. Denver, Colorado: Association for Computational Linguistics, 39–48 [DOI 10.3115/v1/W15-1506](https://doi.org/10.3115/v1/W15-1506).
- Scarselli F, Gori M, Tsoi AC, Hagenbuchner M, Monfardini G. 2009.** The graph neural network model. *IEEE Transactions on Neural Networks* 20:61–80 [DOI 10.1109/TNN.2008.2005605](https://doi.org/10.1109/TNN.2008.2005605).
- Socher R, Huval B, Manning CD, Ng AY. 2012.** Semantic compositionality through recursive matrix-vector spaces. In: *Proceedings of the 2012 joint conference on*

- empirical methods in natural language processing and computational natural language learning*. Jeju Island, Korea: Association for Computational Linguistics, 1201–1211.
- Tai KS, Socher R, Manning CD. 2015.** Improved semantic representations from tree-structured long short-term memory networks. In: *Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing (volume 1: long papers)*. Beijing, China: Association for Computational Linguistics, 1556–1566 DOI [10.3115/v1/P15-1150](https://doi.org/10.3115/v1/P15-1150).
- Wu S, He Y. 2019.** Enriching pre-trained language model with entity information for relation classification. In: *Proceedings of the 28th ACM international conference on information and knowledge management*.
- Xu C, Yuan L, Zhong Y. 2020.** Chinese relation extraction using lattice GRU. In: *2020 IEEE 4th information technology, networking, electronic and automation control conference (ITNEC), vol. 1*. 1188–1192 DOI [10.1109/ITNEC48623.2020.9085019](https://doi.org/10.1109/ITNEC48623.2020.9085019).
- Xue K, Zhou Y, Ma Z, Ruan T, Zhang H, He P. 2019.** Fine-tuning BERT for joint entity and relation extraction in chinese medical text. In: *2019 IEEE international conference on bioinformatics and biomedicine (BIBM)*. Piscataway: IEEE Computer Society, 892–897 DOI [10.1109/BIBM47256.2019.8983370](https://doi.org/10.1109/BIBM47256.2019.8983370).
- Zeng D, Liu K, Chen Y, Zhao J. 2015.** Distant supervision for relation extraction via piecewise convolutional neural networks. In: *Proceedings of the 2015 conference on empirical methods in natural language processing*. Lisbon: Association for Computational Linguistics, 1753–1762 DOI [10.18653/v1/D15-1203](https://doi.org/10.18653/v1/D15-1203).
- Zeng D, Liu K, Lai S, Zhou G, Zhao J. 2014.** Relation classification via convolutional deep neural network. In: *Proceedings of COLING 2014, the 25th international conference on computational linguistics: technical papers*. Dublin: Dublin City University and Association for Computational Linguistics, 2335–2344.
- Zhang D, Wang D. 2015.** Relation classification via recurrent neural network. ArXiv preprint. [arXiv:1508.01006](https://arxiv.org/abs/1508.01006).
- Zhang Y, Qi P, Manning CD. 2018.** Graph convolution over pruned dependency trees improves relation extraction. In: *Proceedings of the 2018 conference on empirical methods in natural language processing*. Brussels: Association for Computational Linguistics, 2205–2215 DOI [10.18653/v1/D18-1244](https://doi.org/10.18653/v1/D18-1244).
- Zhang Y, Yang J. 2018.** Chinese NER using lattice LSTM. In: *Proceedings of the 56th annual meeting of the association for computational linguistics (volume 1: long papers)*. Melbourne: Association for Computational Linguistics, 1554–1564 DOI [10.18653/v1/P18-1144](https://doi.org/10.18653/v1/P18-1144).
- Zhang Y, Zhong V, Chen D, Angeli G, Manning CD. 2017.** Position-aware attention and supervised data improve slot filling. In: *Proceedings of the 2017 conference on empirical methods in natural language processing*. Copenhagen: Association for Computational Linguistics, 35–45 DOI [10.18653/v1/D17-1004](https://doi.org/10.18653/v1/D17-1004).
- Zhong Z. 2021.** Chinese entity relation extraction based on multi-level gated recurrent mechanism and self-attention. In: *2021 2nd international conference on artificial*

intelligence and information systems, ICAIIS 2021. New York: Association for Computing Machinery DOI [10.1145/3469213.3471335](https://doi.org/10.1145/3469213.3471335).

Zhou W, Chen M. 2022. An improved baseline for sentence-level relation extraction. In: *Proceedings of the 2nd conference of the Asia-Pacific chapter of the association for computational linguistics and the 12th international joint conference on natural language processing (Volume 2: Short Papers)*. Association for Computational Linguistics, 161–168.

Zhou P, Shi W, Tian J, Qi Z, Li B, Hao H, Xu B. 2016. Attention-based bidirectional long short-term memory networks for relation classification. In: *Proceedings of the 54th annual meeting of the association for computational linguistics (volume 2: short papers)*. Berlin: Association for Computational Linguistics, 207–212 DOI [10.18653/v1/P16-2034](https://doi.org/10.18653/v1/P16-2034).

Zhu L, Xu M, Bao Y, Xu Y, Kong X. 2022. Deep learning for aspect-based sentiment analysis: a review. *PeerJ Computer Science* **8**:e1044 DOI [10.7717/peerj-cs.1044](https://doi.org/10.7717/peerj-cs.1044).

Zhu L, Zhu Z, Zhang C, Xu Y, Kong X. 2023. Multimodal sentiment analysis based on fusion methods: a survey. *Information Fusion* **95**:306–325 DOI [10.1016/j.inffus.2023.02.028](https://doi.org/10.1016/j.inffus.2023.02.028).