

Improving Relation Classification by Entity Pair Graph

Yi Zhao
Huaiyu Wan
Jianwei Gao
Youfang Lin

YIZHAO@BJTU.EDU.CN
HYWAN@BJTU.EDU.CN
GAOJIANWEI@BJTU.EDU.CN
YFLIN@BJTU.EDU.CN

School of Computer and Information Technology, Beijing Jiaotong University, Beijing, China
Beijing Key Laboratory of Traffic Data Analysis and Mining, Beijing, China

Editors: Wee Sun Lee and Taiji Suzuki

Abstract

Relation classification is one of the most important tasks in the field of information extraction, and also a key component of systems that require relational understanding of unstructured text. Existing relation classification approaches mainly rely on exploiting external resources and background knowledge to improve the performance and ignore the correlations between entity pairs which are helpful for relation classification. We present the concept of entity pair graph to represent the correlations between entity pairs and propose a novel entity pair graph based neural network (EPGNN) model, relying on graph convolutional network to capture the topological features of an entity pair graph. EPGNN combines sentence semantic features generated by pre-trained BERT model with graph topological features for relation classification. Our proposed model makes full use of a given corpus and forgoes the need of external resources and background knowledge. The experimental results on two widely used dataset: SemEval 2010 Task 8 and ACE 2005, show that our method outperforms the state-of-the-art approaches.

Keywords: relation classification, entity pair graph, graph convolutional network, topological structure

1. Introduction

Relation classification is the task of identifying the semantic relation between pairs of entities in given sentences, which is one of the most important tasks in the field of information extraction. For the following sentence

”Huge [*money*] _{e_1} is given to [*companies*] _{e_2} for boosting economy.”

with annotated target entity mentions $e_1 = \text{”money”}$ and $e_2 = \text{”companies”}$, the aim of the relation classification task is to recognize the Entity-Destination relationship between e_1 and e_2 . Relation classification plays a vital role in natural language processing (NLP) applications that require relational understanding of unstructured text, e.g., question answering applications, knowledge base population and knowledge graph construction [Yu et al. (2017); Zhang et al. (2017)]. Many high level NLP tasks, such as precise sentence interpretation and discourse processing [Hendrickx et al. (2010)], can be benefited from accurate relation classification. Thus, relation classification has attracted considerable attention

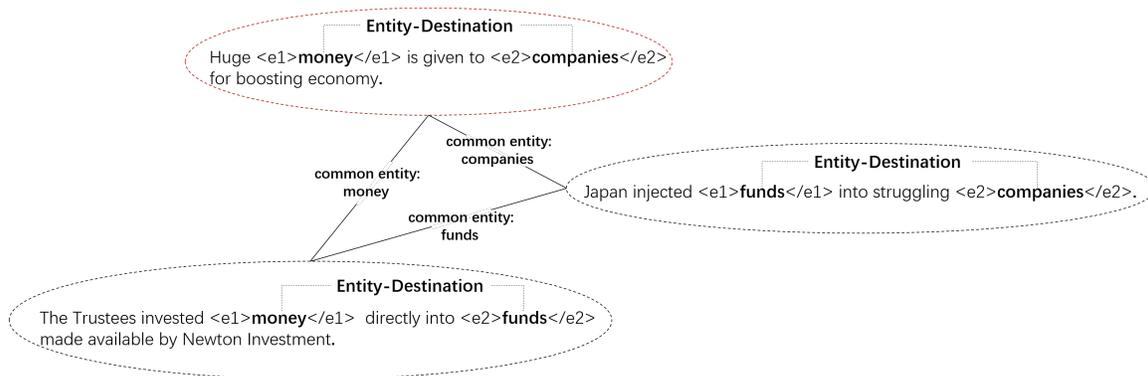


Figure 1: Illustration of the correlation between entity pairs in SemEval 10 Task 8 dataset. The target entity pair is marked with red circle and its correlated entity pairs are marked with black circles. The words on edges are shared entity mentions of linked examples.

from researchers over the course of the past decades [Zhang (2004); Qian et al. (2009); Rink and Harabagiu (2010)].

The field of relation classification has been greatly benefited from recent developments in neural network models, which are capable of learning relevant representations and features without extensive manual feature engineering. A number of neural network architectures have been proposed for relation classification, such as convolutional neural networks (CNN) and recurrent neural networks (RNN) [Zeng et al. (2014); dos Santos et al. (2015); Xu et al. (2015)]. However, those neural models, which only extract semantic features from the given sentences, often fail to outperform feature based and kernel based methods. It is not surprising that researchers tend to take advantage of external resources and background knowledge to provide more information that could be helpful to this task [Ji et al. (2017); Miwa and Bansal (2016)]. Still, these models suffer from several drawbacks. Specifically, the models using background knowledge can be too limited to certain corpus, because background knowledge are often in different forms and sometimes we are even not able to find appropriate knowledge. Similarly, since external resources are always generated by NLP tools, models making use of external resources, such as dependency parsing based models, may suffer from the error propagation and accumulation.

Most existing relation classification models treat each pair of entities individually [Miwa and Bansal (2016); Nguyen and Grishman (2015a)]. However, the relation between a pair of entities can be indicated by other pairs which contain the common entity mentions. Relation classification models could take these entity pairs into consideration to model the dependencies among them. The example illustrated in Fig. 1 explains this phenomenon. The relation between target entity mentions *money* and *companies* can be extracted directly from the target entities or indirectly by incorporating information from its related pairs $\langle money, funds \rangle$ and $\langle funds, companies \rangle$. Intuitively, we can deduce that the relation of target entity pair $\langle money, companies \rangle$ is Entity-Destination from the information that both

$\langle money, funds \rangle$ and $\langle funds, companies \rangle$ have the Entity-Destination relation. This kind of correlations between entity pairs can be represented as an entity pair graph, where entity pairs constitute the nodes and an edge links two entity pairs which contain a common entity mention. It will be of great value to make use of this entity pair graph in the relation classification task.

In this work, we propose a novel entity pair graph based neural network (EPGNN) that is tailored for relation classification to address the mentioned shortcomings of the previous approaches. Instead of relying on the external resources and background knowledge to promote the performance of relation classification, we take advantage of the entity pair graph. In our model, we apply the pre-trained BERT model [Devlin et al. (2019)] to encode contextual information. Meanwhile, we exploit the topological information over the entity pair graph with efficient graph convolutional operation. Finally we combine the topological features of the target entity pair with the semantic features of the input sentence to make robust relation classification.

The main contributions of this paper can be summarized as follows:

1. We present the concept of entity pair graph and propose a graph neural network model for relation classification, which is able to incorporate sentence semantic features and graph topological features for relation classification.
2. The topological feature extractor of our model can be used as a plug-and-play technique to improve the performance for other relation classification models.
3. We achieve the new state-of-the-art results of relation classification on both the SemEval 2010 Task 8 dataset and the ACE 2005 dataset without any use of external resources and background knowledge.

2. Related Work

Most existing relation classification models treat this task as a supervised multi-class classification problem, except a few unsupervised clustering methods. Traditional feature based methods get high performance through a variety of elaborately designed features which are obtained from the external NLP tools, such as POS, syntactic parsing and dependency parsing [Kambhatla (2004); Suchanek et al. (2006); Nguyen and Grishman (2014)]. Various kernel based approaches also leverage syntactic information to measure similarity between two data samples to predict the relation, finding that tree based kernels [Qian et al. (2008)] and subsequence kernels [Bunescu and Mooney (2005)] are effective for this work. Since these methods strongly depend on external NLP tools and lexical and semantic resources, which are not infallible, their performance may be limited by potential errors of the external resources. And because NLP tools always get trained on encyclopedia corpus, these models may lack robustness when applied to new domains.

Recent studies have found deep neural models effective in relation classification. Comparing with the traditional relation classification methods, deep neural networks can learn underlying semantic features automatically, which reduce the dependency on the preprocessing and greatly weaken the negative impact of NLP tools. Socher et al. (2012) employed RNN to learn a representation of the path between entities in syntactic parse tree

and utilized this representation for relation classification. Zeng et al. (2014) first applied CNN with manual features to encode relations. The Ranking-CNN model proposed by dos Santos et al. (2015) can reduce the dependency on manually crafted features and greatly promote the performance by devising a new objective function. Miwa and Bansal (2016) found that LSTM based model which can capture linguistic and syntactic properties of long word sequences can get better performance than CNN based models.

Attention mechanism has been employed to capture the most helpful information for this task. Zhou et al. (2016) and Wang et al. (2016) proposed to use attention mechanism over RNN and CNN architectures for this work. Ji et al. (2017) introduced the entity descriptions from Freebase and Wikipedia as the background knowledge and proposed a sentence-level attention model to select the valid training instances, which achieved a good performance. Nguyen and Grishman (2015b) proposed to combine the traditional feature based method, the convolutional and recurrent neural network to simultaneously benefit from their edges.

Incorporating dependency trees into neural models has also been shown to improve relation classification performance by capturing long distance relation. Xu et al. (2015) generalized the idea of dependency path kernels by applying a LSTM network over the shortest dependency path between entities. Miwa and Bansal (2016) applied a Tree-LSTM [Tai et al. (2015)], a generalized form of LSTM over dependency trees, in a joint entity and relation extraction setting.

In recent years, graph based models [Gong et al. (2015); Gong et al. (2017); Gong et al. (2019)] are applied to relation classification and achieved the solid results. Zhang et al. (2018) first applied graph convolutional network (GCN) [Kipf and Welling (2017)] to relation classification. Their model can pool information over arbitrary dependency structures efficiently in parallel. Christopoulou et al. (2019) proposed a graph based neural network model which treated multiple pairs in sentence simultaneously.

From above all, we can see that neural network based methods have been proved to be effective on relation classification, especially graph based methods. And most of previously proposed approaches rely on external resources and background knowledge. Although these methods work efficiently, we could mine further helpful information from given corpus to improve the performance on relation classification. The rest part of this paper demonstrate that we can get rid of external resources and background knowledge without any degradation of performance.

3. Proposed Approach

3.1. Definition and Notation

Relation Classification. Given a corpus $\mathcal{C} = \{s_1, s_2, \dots, s_m\}$ consisting of m sentences, for one sentence in $s = \{w_1, w_2, \dots, w_n\}$ in \mathcal{C} , w_i indicates the i -th word in the sentence. We use e_1 and e_2 to represent the two entities appearing in sentence s . The task of relation classification aims at predicting the relation of entity pair $\langle e_1, e_2 \rangle$ in sentence s .

Entity Pair Graph. We propose the concept of entity pair graph in this paper to represent the correlations between entity pairs which could contribute to relation classification, where one node indicates an entity pair and an edge connects two entity pairs which contain a common entity. Given a corpus \mathcal{C} , its entity pair graph can be represented as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$,

Notation	Definition
\mathcal{G}	entity pair graph
H_i	final hidden state of BERT
f^{head}	head entity features
f^{tail}	tail entity features
f^s	sentence semantic features
f^g	graph topological features
\tilde{A}	adjacency matrix
\tilde{D}	degree matrix

Table 1: Overview of main notations.

where \mathcal{V} is the set of entity pairs, and each edge $(v_i, v_j) \in \mathcal{E}$ ($v_i, v_j \in \mathcal{V}$) corresponds to a common entity mention between two nodes v_i and v_j .

Table 1 list the notations we will use in this paper.

3.2. Overall Architecture of the Proposed Model

In order to extract a fund of information only from the given corpus, we propose a graph based model which takes entity pairs that contain common entity mentions with target pair into consideration. The overview of our architecture is shown in Fig. 2. For the purpose of encoding contextual information, the input sentence is fed to the BERT model to generate the representations of the whole sentence and the target entities. To obtain the semantic features of the input sentence, a fully connected layer is applied to the concatenation of the generated representations of the input sentence and the target entities. Then the generated vectors of the two target entities are used to compose the representations of \mathcal{V} , in other words, the graph signal of \mathcal{G} . A multi-layer GCN is employed to capture the topological features of \mathcal{G} by calculating new representations of \mathcal{V} . Then, the combination of the sentence semantic features and the graph topological features is used for the final relation classification. The remainder of this section will provide further details about our proposed architecture.

3.3. Sentence and Entity Representation

Given a sentence $S = \{w_1, w_2, \dots, w_n\}$ with two target entities e_1 and e_2 , we use pre-trained BERT model to generate sentence representation and transform every word into a real-valued vector to provide lexical and semantic features.

The BERT model is a pre-trained language model, whose architecture is a multi-layer bidirectional Transformer encoder [Vaswani et al. (2017)]. It is designed to pretrain deep bidirectional representations by jointly conditioning on both left and right context based on BooksCorpus and Wikipedia. Therefore, the pre-trained BERT representations contain much more lexical and semantic information than the word vectors generated by traditional methods which use unidirectional language models and are only based on task specific corpus.

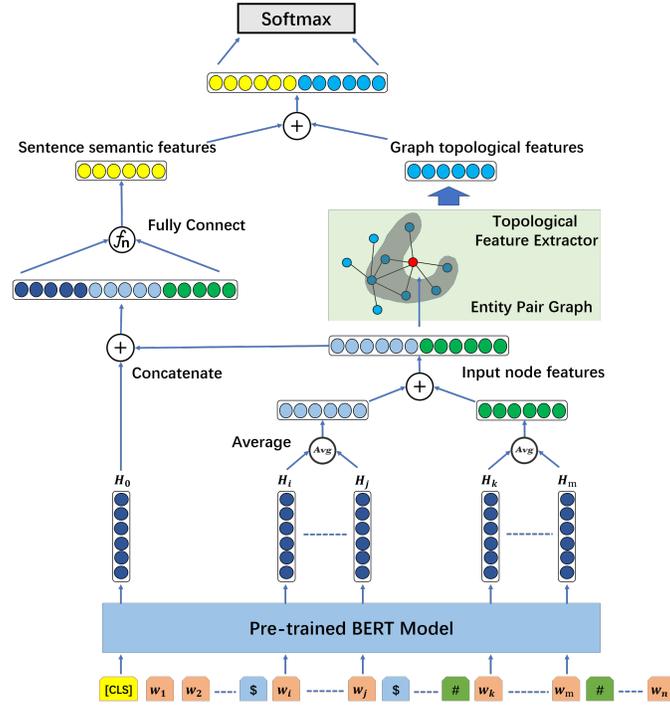


Figure 2: Overall architecture of the EPGNN model.

The input representation of BERT of each token is constructed by summing the corresponding token, segment and position embedding. And a special classification embedding ('[CLS]') is appended to the beginning of each sequence. The output of the Transformer corresponding to this token is used as the aggregate sequence representation for classification. For the purpose of encoding the sentence in an entity-aware manner, we follow [Wu and He \(2019\)](#) to insert a special token '\$' at both the beginning and the end of the head entity. As to the tail entity, we use '#' to indicate its position. For example, the given sequence in Section 1 need to be changed to the following form before fed to the BERT model.

"[CLS] Huge \$ money \$ is given to # companies # for boosting economy."

Entity Representation. For the input sequence S , its final hidden state output from BERT is $\mathcal{H} \in \mathbb{R}^{(n+1) \times d_w}$, where n is the number of words in the input sentence and d_w is the number of dimensions of word vectors determined by the BERT model. Since an entity may be made up of multiple words, we suppose that the head entity starts at w_i and ends at w_j . Therefore, the vector of the head entity can be represented as the average of the final hidden state vectors from H_i to H_j . Similarly, the vector of the tail entity consists of the final hidden state vectors from H_k to H_m . Then we apply a fully connected layer to the both entity vectors to map them to the same dimension d_w and obtain the representations of the two target entities, denoted as $f^{head} \in \mathbb{R}^{d_w}$ and $f^{tail} \in \mathbb{R}^{d_w}$. This process can be

mathematically formalized as Equation (1) and (2).

$$f^{head} = W_1 \left[\frac{\sum_{t=i}^j H_t}{j-i+1} \right]^\top + b_1 \quad (1)$$

$$f^{tail} = W_2 \left[\frac{\sum_{t=k}^m H_t}{m-k+1} \right]^\top + b_2 \quad (2)$$

where $W_1 \in \mathbb{R}^{d_w \times d_w}$ and $W_2 \in \mathbb{R}^{d_w \times d_w}$ are weight matrixes, $b_1 \in \mathbb{R}^{d_w}$ and $b_2 \in \mathbb{R}^{d_w}$ are biases.

Sentence Representation. We follow the operation in Devlin et al. (2019) and use the hidden state vector of the first token H_0 as the initial sentence representation, denoted as $f^{cls} \in \mathbb{R}^{d_w}$. We can learn from previous work [Ji et al. (2017)] that the information of target entities plays a crucial role in relation classification. In order to emphasize the importance of entity pairs, we apply a fully connected layer which is followed by a activation, e.g., $ReLU(\cdot)$, to the concatenation of f^{cls} , f^{head} and f^{tail} to represent sentence semantic features $f^s \in \mathbb{R}^{d_s}$, where d_s is the number of dimensions of the sentence semantic vector, which is a hyperparameter. f^s can be mathematically represented as follow:

$$f^s = \sigma(W_3[f^{cls}; f^{head}; f^{tail}]^\top + b_3) \quad (3)$$

where $W_3 \in \mathbb{R}^{d_s \times 3d_w}$ is weight matrix, $b_3 \in \mathbb{R}^{d_s}$ is bias.

3.4. Topological Feature Extractor

As demonstrated in Section 1, the topological features of the entity pair graph can help extract more accurate relations. For the purpose of exploiting the correlation structure over the entity pair graph, our EPGNN model employs a multi-layer GCN over the entity pair graph to encode the graph topological features.

GCNs are neural networks that are modified to work directly on a graph structure. It ingeniously integrate local vertex features and graph topological features. Li et al. (2018) showed that the GCN model is simply a special form of Laplacian smoothing, which mixes the features of a vertex and its nearby neighbors. Therefore, after the smoothing operation, the features of vertices in the same cluster will become more similar. Obviously, it will make classification more accurate.

Adjacency and Degree Matrix. Given an entity pair graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with n nodes, it can be represented with an adjacency matrix $A = [a_{ij}] \in \mathbb{R}^{n \times n}$ where $A_{ij} = 1$ and $A_{ji} = 1$ if there is an edge between node i and node j . In order to take the target entity pair itself into consideration, we add self-connections to the adjacency matrix. So we obtain new adjacency matrix $\tilde{A} = A + I_N$ where I_N is the identity matrix. And the degree matrix of \mathcal{G} is calculated according to \tilde{A} , denoted as $\tilde{D} = \text{diag}(\tilde{d}_1, \tilde{d}_2, \dots, \tilde{d}_n)$ where $\tilde{d}_i = \sum_j \tilde{a}_{ij}$ is the degree of vertex i . After all these calculation based on \mathcal{G} , we get the adjacency matrix \tilde{A} and degree matrix \tilde{D} .

Convolutional Operation on Graph. There are two mainstream implement methods of GCNs: spatial method and spectral method. We employ the spectral GCNs in this work. Spectral GCNs define the convolution by decomposing a graph signal on the spectral domain and then applying a spectral filter on it. According to Kipf and Welling (2017), to

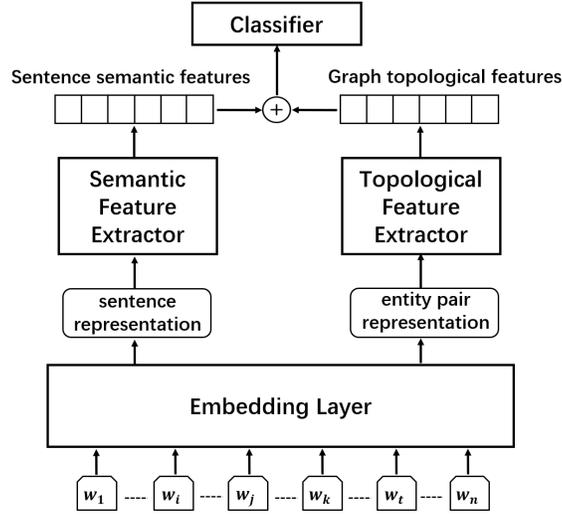


Figure 3: The framework of entity pair graph enhanced relation classifier.

a graph signal $X \in \mathbb{R}^{n \times c}$, where c is the number of dimensions of vertex feature vectors, the spectral convolutional operation with f spectral filters can be mathematically formalized as follow:

$$H^{(l+1)} = \sigma(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} \Theta^{(l)}) \quad (4)$$

where $H^{(l)}$ is the matrix of activation in the l -th layer, and $H^{(0)} = X$, $\Theta \in \mathbb{R}^{c \times f}$ is the trainable weight matrix in layer l , σ is the activation function, e.g., $ReLU(\cdot)$.

In the GCN module of our model, we use the concatenation of f^{head} and f^{tail} to represent a vertex. And the representations of all the vertices form the graph signal $X \in \mathbb{R}^{n \times 2d_w}$. We apply a three-layer GCN to generate the graph topological feature $f^g \in \mathbb{R}^{d_g}$:

$$f^g = \hat{A} \sigma(\hat{A} \sigma(\hat{A} X W^{(0)}) W^{(1)}) W^{(2)} \quad (5)$$

where $\hat{A} = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}}$, $W^{(0)} \in \mathbb{R}^{2d_w \times d_h}$ and $W^{(1)} \in \mathbb{R}^{d_h \times d_h}$ are weight matrixes for hidden layers with d_h feature maps, $W^{(2)} \in \mathbb{R}^{d_h \times d_g}$ is a hidden-to-output weight matrix, d_g is the number of dimensions of the graph topological features. d_h and d_g are hyperparameters.

Plug-and-play Technique. As depicted in Fig. 3, the topological feature extractor of our model can be applied to other relation classification models and improve their performance without any change of the original feature extractors. To achieve this, we only need two small modifications. Firstly, the representation of the target entity pair need to be output from the embedding layer and fed into the topological feature extractor to generate graph topological features. Secondly, we use the concatenation of the original features and the graph topological features to make final classification. Our experiments in Section 4 demonstrate that adding the topological feature extractor can indeed improve the performance of relation classifiers.

3.5. Relation Classification

After all above operations, we use a softmax-classifier to predict the relation of the entity pair $\langle e_1, e_2 \rangle$. The input of the softmax-classifier is the concatenation of the sentence representation $f^s \in \mathbb{R}^{d_s}$ and the target entity pair representation $f^g \in \mathbb{R}^{d_g}$:

$$p(y|s) = \text{softmax}(W_0[f^s; f^g]^\top + b_0) \quad (6)$$

where $W_0 \in \mathbb{R}^{r \times (d_s + d_g)}$ is a weight matrix, $b_0 \in \mathbb{R}^r$ is a bias and r is a hyperparameter equaling to the number of relation types.

Then we can get the predicted label of the target entity pair $\langle e_1, e_2 \rangle$ according to the distribution $p(y|s)$:

$$\hat{y} = \arg \max_y p(y|s) \quad (7)$$

The loss function of the EPGNN model is the cross entropy:

$$\mathcal{L} = - \sum_i \sum_j Y_j \log p(y_{ij}|s_i, \theta) \quad (8)$$

where θ indicates all the parameters in the model.

In order to compute the network parameter θ , we minimize the cross entropy loss using Adam. The parameters of the network are randomly initialized and the back propagation algorithm is employed to update them.

4. Experiments

4.1. Dataset and Metric

We conduct our experiments on two widely used relation classification datasets: (1) **SemEval 2010 Task 8**: The dataset contains 10,717 examples annotated with 9 different relation types and an additional "Other" type, which is used to indicate that the relation in the sentence is not among the 9 types. For the relation is directional, the sentence needs to be regarded as expressing a different relation if the target entities appear in inverse order. For example, the relation Cause-Effect(e_1, e_2) and Cause-Effect(e_2, e_1) are considered as two distinct relations. The SemEval dataset is already partitioned into 8,000 training examples and 2,717 test examples. We evaluate the models using the official scorer in terms of the Macro-F1 score over the 9 relation types (excluding Other) and take the directionality into consideration. (2) **ACE 2005**: In this dataset, 7123 examples in 511 documents are annotated with 6 major relation types. We use the same data splits, preprocessing, and task settings as [Christopoulou et al. \(2019\)](#). We report the primary Micro-F1 scores as evaluation metrics.

4.2. Experimental Settings

We use the $BERT_{BASE}$ model [[Devlin et al. \(2019\)](#)] to encode input sentences and target entities. In order to avoid overfitting during the training process, we adopt the dropout strategy in classification components. We tune all the hyperparameters by using cross-validation procedure on the training sets. The choices generated by this process are given in Table 2.

Parameter	Parameter Name	Value
d_s	sentence feature dimension	768
d_h	number of feature maps in hidden layer of GCN	128
d_g	topological feature dimension	64
l_{max}	max sentence length	128
$batch_size$	batch size	16
dp	dropout rate	0.3
λ	learning rate	5e-6

Table 2: Hyperparameter setting of the EPGNN model on SemEval 10.

4.3. Results on the SemEval 2010 Dataset

On the SemEval dataset, our EPGNN model outperforms the previously proposed models with an Macro-F1 score of 90.2%. Table 3 provides a detailed comparison of our EPGNN model with previous approaches. Dependency methods are neural network models which utilize a grammar parser to predict relation classes. We refer to other neural network models those work without using a grammar parser as end-to-end methods. As to methods using pre-trained language models, we name them pre-trained methods.

We observe that our proposed model achieves new state-of-the-art results on this relation classification dataset. Our EPGNN outperforms the approaches which rely on external resources such as dependency parser, POS and named entity tags [Xu et al. (2015); Miwa and Bansal (2016); Zhang et al. (2018)]. This indicates that our model can indeed mine a fund of information from the given corpus to make accurate relation classification without any use of external resources. Because of this property, our model is more applicable to the scenario where appropriate external resources are unacquirable. In addition, our EPGNN model gets higher Macro-F1 score than the TRE and R-BERT model which employ pre-trained language model to make relation classification. Outperforming these methods highlights the effectiveness of adding graph topological features for relation classification, which we will make further discussion in Section 4.5.

4.4. Results on the ACE 2005 Dataset

On the ACE dataset, We follow the preprocessing process of Christopoulou et al. (2019) and compare the performance with their walk based model which achieved the stat-of-the-art results on the ACE 2005 dataset so far. In order to demonstrate the effectiveness of the topological feature extractor, we retrain the R-BERT model to make comparison. Table 4 illustrates the performance of the EPGNN model in comparison with the state-of-the-art walk based method and the R-BERT model. Our EPGNN model outperforms the walk based model with a relative improvement of 4.2%. The walk based model treats multiple pairs in a sentence simultaneously and consider the interactions among them. Outperforming this approach illustrates that interactions among entity pairs can truly help relation classification and the corpus-level correlations between entity pairs are more efficient than the sentence-level correlations. And comparing with the R-BERT model, EPGNN still makes

Model	Macro-F1
<i>Manually Engineered Methods</i>	
SVM [Rink and Harabagiu (2010)]	82.2
<i>Dependency Methods</i>	
FCM [Yu et al. (2014)]	83.0
SDP-LSTM [Xu et al. (2015)]	83.7
DRNNs [Xu et al. (2016)]	85.8
SPTree [Miwa and Bansal (2016)]	84.5
C-GCN [Zhang et al. (2018)]	84.8
<i>End-to-End Methods</i>	
CNN+Softmax [Zeng et al. (2014)]	82.7
CR-CNN [dos Santos et al. (2015)]	84.1
Att-Pooling-CNN [Wang et al. (2016)]	88.0
<i>Pre-Trained Methods</i>	
TRE [Alt et al. (2019)]	87.1
R-BERT [Wu and He (2019)]	89.2
<i>Our Method</i>	
EPGNN	90.2

Table 3: Comparison with results on SemEval 10 published in the literature.

nearly 1% improvement on Micro-F1 score. This result signals that EPGNN indeed benefit from the topological feature extractor.

Model	Micro-F1
walk based model [Christopoulou et al. (2019)]	72.9
R-BERT [Wu and He (2019)]	76.2
EPGNN	77.1

Table 4: Comparison with results on ACE 2005 published in the literature.

4.5. Ablation Study

The experimental results have demonstrated the effectiveness of the topological feature extractor in our proposed model. Since the topological feature extractor is an independent component, we consider to apply it to other relation classification models and perform ablation experiments to demonstrate that it can be used as an effective plug-and-play technique. We conduct experiments on the SemEval dataset with three representative neural network models: CNN based model, LSTM based model and BERT based model. The results of the ablation experiments are provided in Table 5. The first two models make more than 2% improvement on performance by the topological feature extractor. For the BERT model that already achieves a relatively high Macro-F1 score, adding topological features still

Model	Macro-F1
<i>CNN based Methods</i>	
without GCN module	75.3
with GCN module	77.6
<i>LSTM based Methods</i>	
without GCN module	79.6
with GCN module	82.1
<i>BERT based Methods</i>	
without GCN module	89.2
with GCN module	90.2

Table 5: Ablation with and without the GCN module on SemEval 10.

makes 1% improvement. The results show that adding topological features can improve the performance regardless of the original architecture. The kernel components of the topological feature extractor is a multi-layer GCN. As explained in Section 3.4, the GCN model is a special form of Laplacian smoothing. It can compute new features of a vertex as the weighted average of itself and its neighbors. Therefore, this operation makes connected nodes more similar and thus easier for classification. This kind of topological information of vertices was ignored in previously proposed approaches. Hence, adding the topological information of the entity pair graph is of great value of to relation classification.

5. Conclusion

In this paper, we propose the concept of entity pair graph whose topological structure implies the correlations between entity pairs. And we present a novel entity pair graph based neural network model EPGNN which enriches the BERT based relation classification model with the topological information extracted from the entity pair graph. We obtain new state-of-the-art F1 scores on two popular relation classification datasets: SemEval 2010 Task 8 and ACE 2005. The results show that the topological features of the entity pair graph can indeed help the relation classification task. And the performance of EPGNN shows that models without any use of external resources and background knowledge can also achieve a competitive results on this task. We apply the topological feature extractor of EPGNN to some previous relation classification models and achieve better performances, which means that it can be used as a plug-and-play technique. The important future work would be to modify our model to address the wrong label problem existing in the distant supervision dataset.

Acknowledgments

This paper is supported by the National Key R&D Program of China (No. 2018YFC0830200). We thank anonymous reviewers for their valuable comments

References

- Christoph Alt, Marc Hübner, and Leonhard Hennig. Improving relation extraction by pre-trained language representations. *CoRR*, abs/1906.03088, 2019. URL <http://arxiv.org/abs/1906.03088>.
- Razvan C. Bunescu and Raymond J. Mooney. Subsequence kernels for relation extraction. In *Advances in Neural Information Processing Systems 18 [Neural Information Processing Systems, NIPS 2005, December 5-8, 2005, Vancouver, British Columbia, Canada]*, pages 171–178, 2005. URL <http://papers.nips.cc/paper/2787-subsequence-kernels-for-relation-extraction>.
- Fenia Christopoulou, Makoto Miwa, and Sophia Ananiadou. A walk-based model on entity graphs for relation extraction. *CoRR*, abs/1902.07023, 2019. URL <http://arxiv.org/abs/1902.07023>.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 4171–4186, 2019. URL <https://aclweb.org/anthology/papers/N/N19/N19-1423/>.
- Cícero Nogueira dos Santos, Bing Xiang, and Bowen Zhou. Classifying relations by ranking with convolutional neural networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL 2015, July 26-31, 2015, Beijing, China, Volume 1: Long Papers*, pages 626–634, 2015. URL <http://aclweb.org/anthology/P/P15/P15-1061.pdf>.
- Chen Gong, Tongliang Liu, Dacheng Tao, Keren Fu, Enmei Tu, and Jie Yang. Deformed graph laplacian for semisupervised learning. *IEEE Trans. Neural Netw. Learning Syst.*, 26(10):2261–2274, 2015. doi: 10.1109/TNNLS.2014.2376936. URL <https://doi.org/10.1109/TNNLS.2014.2376936>.
- Chen Gong, Dacheng Tao, Wei Liu, Liu Liu, and Jie Yang. Label propagation via teaching-to-learn and learning-to-teach. *IEEE Trans. Neural Netw. Learning Syst.*, 28(6):1452–1465, 2017. doi: 10.1109/TNNLS.2016.2514360. URL <https://doi.org/10.1109/TNNLS.2016.2514360>.
- Chen Gong, Dacheng Tao, Xiaojun Chang, and Jian Yang. Ensemble teaching for hybrid label propagation. *IEEE Trans. Cybernetics*, 49(2):388–402, 2019. doi: 10.1109/TCYB.2017.2773562. URL <https://doi.org/10.1109/TCYB.2017.2773562>.
- Iris Hendrickx, Su Nam Kim, Zornitsa Kozareva, Preslav Nakov, Diarmuid Ó Séaghdha, Sebastian Padó, Marco Pennacchiotti, Lorenza Romano, and Stan Szpakowicz. Semeval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals. In *Proceedings of the 5th International Workshop on Semantic Evaluation, SemEval@ACL*

- 2010, Uppsala University, Uppsala, Sweden, July 15-16, 2010, pages 33–38, 2010. URL <http://aclweb.org/anthology/S/S10/S10-1006.pdf>.
- Guoliang Ji, Kang Liu, Shizhu He, and Jun Zhao. Distant supervision for relation extraction with sentence-level attention and entity descriptions. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA.*, pages 3060–3066, 2017. URL <http://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14491>.
- Nanda Kambhatla. Combining lexical, syntactic, and semantic features with maximum entropy models for information extraction. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics, Barcelona, Spain, July 21-26, 2004 - Poster and Demonstration*, 2004. URL <https://aclanthology.info/papers/P04-3022/p04-3022>.
- Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*, 2017. URL <https://openreview.net/forum?id=SJU4ayYgl>.
- Qimai Li, Zhichao Han, and Xiao-Ming Wu. Deeper insights into graph convolutional networks for semi-supervised learning. In *AAAI*, 2018.
- Makoto Miwa and Mohit 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*, 2016. URL <http://aclweb.org/anthology/P/P16/P16-1105.pdf>.
- Thien Huu Nguyen and Ralph Grishman. Employing word representations and regularization for domain adaptation of relation extraction. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014, June 22-27, 2014, Baltimore, MD, USA, Volume 2: Short Papers*, pages 68–74, 2014. URL <http://aclweb.org/anthology/P/P14/P14-2012.pdf>.
- Thien Huu Nguyen and Ralph Grishman. Relation extraction: Perspective from convolutional neural networks. In *Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing, VS@NAACL-HLT 2015, June 5, 2015, Denver, Colorado, USA*, pages 39–48, 2015a. URL <http://aclweb.org/anthology/W/W15/W15-1506.pdf>.
- Thien Huu Nguyen and Ralph Grishman. Combining neural networks and log-linear models to improve relation extraction. *CoRR*, abs/1511.05926, 2015b. URL <http://arxiv.org/abs/1511.05926>.
- Longhua Qian, Guodong Zhou, Fang Kong, Qiaoming Zhu, and Peide Qian. Exploiting constituent dependencies for tree kernel-based semantic relation extraction. In *COLING 2008, 22nd International Conference on Computational Linguistics, Proceedings of the Conference, 18-22 August 2008, Manchester, UK*, pages 697–704, 2008. URL <http://www.aclweb.org/anthology/C08-1088>.

- Longhua Qian, Guodong Zhou, Fang Kong, and Qiaoming Zhu. Semi-supervised learning for semantic relation classification using stratified sampling strategy. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, EMNLP 2009, 6-7 August 2009, Singapore, A meeting of SIGDAT, a Special Interest Group of the ACL*, pages 1437–1445, 2009. URL <http://www.aclweb.org/anthology/D09-1149>.
- Bryan Rink and Sanda M. Harabagiu. UTD: classifying semantic relations by combining lexical and semantic resources. In *Proceedings of the 5th International Workshop on Semantic Evaluation, SemEval@ACL 2010, Uppsala University, Uppsala, Sweden, July 15-16, 2010*, pages 256–259, 2010. URL <http://aclweb.org/anthology/S/S10/S10-1057.pdf>.
- Richard Socher, Brody Huval, Christopher D. Manning, and Andrew Y. Ng. 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, EMNLP-CoNLL 2012, July 12-14, 2012, Jeju Island, Korea*, pages 1201–1211, 2012. URL <http://www.aclweb.org/anthology/D12-1110>.
- Fabian M. Suchanek, Georgiana Ifrim, and Gerhard Weikum. Combining linguistic and statistical analysis to extract relations from web documents. In *Proceedings of the Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Philadelphia, PA, USA, August 20-23, 2006*, pages 712–717, 2006. doi: 10.1145/1150402.1150492. URL <https://doi.org/10.1145/1150402.1150492>.
- Kai Sheng Tai, Richard Socher, and Christopher D. Manning. 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 of the Asian Federation of Natural Language Processing, ACL 2015, July 26-31, 2015, Beijing, China, Volume 1: Long Papers*, pages 1556–1566, 2015. URL <http://aclweb.org/anthology/P/P15/P15-1150.pdf>.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*, pages 6000–6010, 2017. URL <http://papers.nips.cc/paper/7181-attention-is-all-you-need>.
- Linlin Wang, Zhu Cao, Gerard de Melo, and Zhiyuan Liu. Relation classification via multi-level attention cnns. 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*, 2016. URL <http://aclweb.org/anthology/P/P16/P16-1123.pdf>.
- Shanchan Wu and Yifan He. Enriching pre-trained language model with entity information for relation classification. *ArXiv*, abs/1905.08284, 2019.
- Yan Xu, Lili Mou, Ge Li, Yunchuan Chen, Hao Peng, and Zhi Jin. Classifying relations via long short term memory networks along shortest dependency paths. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP*

- 2015, Lisbon, Portugal, September 17-21, 2015, pages 1785–1794, 2015. URL <http://aclweb.org/anthology/D/D15/D15-1206.pdf>.
- Yan Xu, Ran Jia, Lili Mou, Ge Li, Yunchuan Chen, Yangyang Lu, and Zhi Jin. Improved relation classification by deep recurrent neural networks with data augmentation. In *COLING 2016, 26th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, December 11-16, 2016, Osaka, Japan*, pages 1461–1470, 2016. URL <http://aclweb.org/anthology/C/C16/C16-1138.pdf>.
- Mo Yu, Matthew Gormley, and Mark Dredze. Factor-based compositional embedding models. In *NIPS Workshop on Learning Semantics*, pages 95–101, 2014.
- Mo Yu, Wenpeng Yin, Kazi Saidul Hasan, Cícero Nogueira dos Santos, Bing Xiang, and Bowen Zhou. Improved neural relation detection for knowledge base question answering. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers*, pages 571–581, 2017. doi: 10.18653/v1/P17-1053. URL <https://doi.org/10.18653/v1/P17-1053>.
- Daojian Zeng, Kang Liu, Siwei Lai, Guangyou Zhou, and Jun Zhao. Relation classification via convolutional deep neural network. In *COLING 2014, 25th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, August 23-29, 2014, Dublin, Ireland*, pages 2335–2344, 2014. URL <http://aclweb.org/anthology/C/C14/C14-1220.pdf>.
- Yuhao Zhang, Victor Zhong, Danqi Chen, Gabor Angeli, and Christopher D. Manning. Position-aware attention and supervised data improve slot filling. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017*, pages 35–45, 2017. URL <https://aclanthology.info/papers/D17-1004/d17-1004>.
- Yuhao Zhang, Peng Qi, and Christopher D. Manning. Graph convolution over pruned dependency trees improves relation extraction. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, pages 2205–2215, 2018. URL <https://aclanthology.info/papers/D18-1244/d18-1244>.
- Zhu Zhang. Weakly-supervised relation classification for information extraction. In *Proceedings of the 2004 ACM CIKM International Conference on Information and Knowledge Management, Washington, DC, USA, November 8-13, 2004*, pages 581–588, 2004. doi: 10.1145/1031171.1031279. URL <https://doi.org/10.1145/1031171.1031279>.
- Peng Zhou, Wei Shi, Jun Tian, Zhenyu Qi, Bingchen Li, Hongwei Hao, and Bo Xu. Attention-based bidirectional long short-term memory networks for relation classification. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 2: Short Papers*, 2016. URL <http://aclweb.org/anthology/P/P16/P16-2034.pdf>.