IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 11 November 2023, accepted 26 November 2023, date of publication 30 November 2023, date of current version 8 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3337871

RESEARCH ARTICLE

DoreBer: Document-Level Relation Extraction Method Based on BernNet

BOYA YUAN AND LIWEN XU
College of Science, North China University of Technology, Beijing 100144, China
Corresponding author: Liwen Xu (xulw163@163.com)

This work was supported in part by the National Social Science Foundation of China under Grant 20BTJ046.

ABSTRACT Document-level relation extraction (RE) task aims to predict predefined relation types of every entity pair in a given document. Compared with the sentence-level counterpart, document-level relation extraction task requires reasoning in a more complex environment, where exist much longer text and much larger amount of entities, making it a more challenging task. However, previous methods suffers from over-smoothing problem when the count of GNN layers is high enough, leading high frequency signals on graph could not pass through filter, and then resulting in an insufficient approximation of the real function and finally causing a defective performance in tasks. To solve this problem, we propose a novel model, called **DoreBer**, for document-level RE task, which can obtain a higher quality of graph representation. Specifically, DoreBer performs estimation of filter over the normalized Laplacian spectrum of a graph by leveraging an order-K Bernstein polynomial approximation, and designs its spectral property by setting the coefficients of the Bernstein basis. Therefore, DoreBer can alleviate the over-smoothing problem to enhance learning ability of model. In addition, DoreBer has a higher interpretability for learned parameters of graph filter. We evaluate DoreBer on the DocRED public document-level RE dataset. **Online** experimental results demonstrate that DoreBer achieves significant performance improvements (2.72 and 2.76 higher on Ign F_1 and F_1 respectively), over the previous state-of-the-art on sequence-based method baseline. DoreBer reveals the potential of BernNet method in document-level relation extraction tasks and sheds light on a path to learn potential representation in high-dimensional data.The source code of this paper can be obtained from https://github.com/factor77/DoreBer/.

INDEX TERMS Natural language processing, document-level relation extraction, graph induction, graph neural network, attention mechanism.

I. INTRODUCTION

Document-level Relation extraction (RE) is aiming to identify relation types for every pair of entities in a given document, and plays an import role in information extraction. RE is widely used to facilitate a lot of downstream foundational tasks of Natural Language Processing (NLP) including knowledge base construction [1], [2], [3], knowledge graph completion [4], [5], [6], information retrieval [7], [8], [9], alignment [10], [11], [12], and question answering [13], [14]. Due to the capability of integrating information within and across multiple sentences of a document and capturing complex interactions between inter-sentence entities, the document-level relation extraction task appeals to many researchers.

Previous works mainly focus on sentence-level RE, which identify relations between entities only from a single sentence [15], [16]. However, sentence-level RE methods suffer from an ineluctable limitation, large amount of relations, such as relations from Wikipedia text, which are reflected through several sentences in real-world applications [17]. Therefore, it is a necessity of extracting relations at the document-level for comprehensively understanding information from text [18], [19], [20].

There exits several core challenges in document-level RE. Firstly, the subject entity and object entity involved in a

The associate editor coordinating the review of this manuscript and approving it for publication was Yiming Tang.

The Great Hall of the People

[0] The *Great Hall* of the People is a state building located at the western edge of *Tiananmen Square* in *Beijing*.
[1] It is used for legislative and ceremonial activities by the government of the ***People's Republic of China*** (***PRC***) and the ruling Communist Party of China.
[2] The *People's Great Hall* functions as the meeting place for the full sessions of the *National People 's Congress* (*NPC*), the ***Chinese*** legislature, which occurs every year during March along with the national session of the Chinese People's Political Consultative Conference (CPPCC), a political advisory body.

Subject: *Tiananmen Square*
Object: *Beijing*
Relation: located in **Supporting Evidence:** 0

Subject: *National People's Congress*
Object: ***PRC***
Relation: country **Supporting Evidence:** 2

Subject: *Tiananmen Square*
Object: ***PRC***
Relation: country **Supporting Evidence:** 0, 2

FIGURE 1. An example selected from the DocRED dataset. Each document in DocRED is annotated with named entity mentions, coreference information, intra- and inter-sentence relations, and supporting evidence. 3 out of the 11 relation instances annotated for this example document are presented. Mentions corresponding to the same named entity are painted in the same colors and other named entity mentions underlined for clarity. Note that mentions of the same subject (e.g., People's Republic of China, PRC and Chinese) are identified as shown in the first relation instance.

relation may emerge from diverse sentences. Hence, under such circumstance, a correct prediction of a relation requires reasoning across multiple sentences. Secondly, a same entity may emerge from diverse sentences in different ways of mention. Therefore, aggregating across multiple sentences helps to extract a better entity representation. Thirdly, a large amount of relations require logic reasoning method to extract them. Specifically, the identification of some relations necessarily relies on the correct prediction of other relations between their head entities and tail entities. Figure 1 shows an example chosen from the document-level dataset DocRED [21]. DocRED is a large-scale human-annotated document-level RE dataset, consisting of 5, 053 documents and covering a wide range of categories with 96 relation types. And we use human annotation part of DocRED, and the statistics of which are shown in Table 1. As Figure 1 shows, it is easy to recognize the intra-sentence relations <*Tiananmen Square, located in the administrative territorial entity, Beijing*>, <*People 's Great Hall, located in the administrative territorial entity, Beijing*>, and <*Communist Party of China, country, PRC*>, since the subject entity and object entity appear in the same sentence. However, it is non-trivial to predict the inter-sentence relations between *Tiananmen Square* and *PRC*, as well as *National People 's Congress* and *PRC*, whose mentions do not appear in the same sentence and have long-distance dependencies. Besides, the identification of these two relation instances also requires logical reasoning. For example, *Tiananmen Square* belongs to

TABLE 1. Statistics of dataset we used.

Setting	Doc.	Rel.	Inst.	Fact
Train	3,053	96	38,269	34,715
Dev	1,000	96	12,332	11,790
Test	1,000	96	12,842	12,101

PRC because *Tiananmen Square* is located in *Beijing*, which belongs to *PRC*.

To tackle the above challenge, most previous works first retrieve the information in the document to obtain a graph, and then aggregate information from this retrieved graph through GCN [22] to learn features. However, the filter used by GCN could be negative. And, However, models based on GCN suffer from over-smoothing problem when number of layers is high enough, leading high frequency signals on graph could not pass through filter, and resulting in an insufficient approximation of the real function and finally a defective performance in tasks. By using BernNet [23], our proposed model is able to solve the over-smoothing problem when setting high number of layers. In BernNet, the number of layers is set as 10. To deal with the problem of features extraction, which is more complex situations comparing with the task of learning filters from signal in images and the task of node classification, we finally choose to apply 100 layers.

In this paper, we propose a Graph Network for document-level RE, called **DoreBer** (**Do**cument-Level **R**elation **E**xtraction method based on **Ber**nNet). It is designed to solve the problems mentioned above directly. Following [24], DoreBer construct three types of nodes for a document-level graph: mention nodes, entity nodes and meta dependency path(MDP) nodes. Then, following Liu [25] and [24], we leverage structured attention [26] and a variant of Kirchhoff's Matrix-Tree Theorem [27], [28] to induce graph structure. Next, we apply BernNet to aggregate information in the induced graph for obtaining new representation of each node. After that, following [24], we perform a iterative refinement on the updated document-level graph, allowing DoreBer to obtain a more informative graph structure for the final relation prediction of any given entity pair.

Our method reveals the potential of BernNet in extracting relation at document-level sequence-based methods and sheds light on a path to RE methods based on pre-trained models.

This paper contributes to the following aspects:

- We propose a novel model which can alleviate the over-smoothing problem by leveraging BernNet to extract graph features to improve performance on the task.
- We evaluate the proposed method on DocRED dataset **online** and results demonstrate that DoreBer achieves the state-of-the-art performance on the benchmark of sequence-based methods for document-level RE.
- We also perform extensive analyses of our proposed model to better illustrate its working mechanism. Specifically, we conduct 3 ablation studies and 1 case

study, and carry out meticulous analyses based on statistics of results and case results.

II. RELATED WORK

Early efforts on relation extraction mainly focused on the sentence level, which predict relation types between entities within a given single sentence by capturing interactions from the input sequence [15], [29], [30], [31], [32] or the dependency tree of the input sequence [33], [34], [35], [36]. However, these methods do not take interactions across mentions into consideration and neglect relations generated by information from multiple sentences [21], [37].

Consequent works [14], [17], [18], [20] begin to push the extraction task into cross-sentence level. Besides the application of discourse structure understanding methods [38], [39], [40], these approaches make use of the dependency graph to capture information of inter-sentence interactions. However, their attention is still confined to a few sentences.

Then, the research attention has been gradually shifted to the whole document level [37], [41], [42], [43], in the biomedical line, but with only a few relations among chemicals were took into consideration. And, sequence-based architectures have also proven to be highly effective for the DocRE task [44], [45], [46]. Specially, [47] leverage CNN and SVM to effectively extract representation. Besides, methods have been proposed to tackle document-level relation extraction [48], [49], [50], [51].

Then, the research attention has been gradually shifted to the whole document level [37], [41], [42], [43], in the biomedical line, but with only a few relations among chemicals were took into consideration. And, sequence-based architectures have also proven to be highly effective for the DocRE task, [45], [47]. Besides, methods have been proposed to tackle document-level relation extraction [48], [49], [50]. Then, the research attention has been gradually shifted to the whole document level [37], [41], [42], [43], in the biomedical line, but with only a few relations among chemicals were took into consideration. And, sequence-based architectures have also proven to be highly effective for the DocRE task [45], [47]. Besides, methods have been proposed to tackle document-level relation extraction [48], [49], [50].

To alleviate that problem, architectures based on graph neural networks (GNNs) [52] have been widely applied. [18] used words and dependency information as nodes and edges respectively to construct document-level graphs. These graphs will be leveraged to extract information of features for each entity pair. Later methods extended this thought by applying diverse GNN-based architectures to extract information of features [17], [36], [37], [43], [53], [54]. In particular, [24] used structured attention mechanism [55] to generate the task-specific document-level graph, which is capable of aggregating information of interactions between entity pairs over multiple sentences. [56] extract features via reasoning from two graphs, one graph for capturing interaction among mentions, and then another for aggregating mentions of the entities.

Unlike previous document-level relation extraction approaches that use GCN [22], GCNN [42] or GAT [57], our DoreBer model leverage BernNet [23] to extract graph features, allowing the model to alleviate the over-smoothing problem for better relational reasoning.

III. OVERVIEW

A. TASK FORMULATION

Document-level relation extraction is defined as a task to predict the pre-defined relation r exits or not between the entity pair (h, t) mentioned in a document $\mathcal{D} = \{S_1, S_2, \ldots, S_n\}$. $S_i = \{w_1, w_2, \ldots, w_m\}$ is the i^{th} sentence in the document $\mathcal{D}$. w_j is the j^{th} word in the sentence S. Entity h or t is composed of a consequence span of words $\{w_1, w_2, \ldots, w_k\}$. $n(= |\mathcal{D}|)$ is the amount of sentence of the document $\mathcal{D}$. $m(= |S_i|)$ is the amount of words in the S_i. $k(= |h|)$ is the amount of words in the entity h. r is belong to a predefined relation type set $\mathcal{R}$. The difficulty of the document-level relation extraction task, comparing with the sentence-level one, is that the entity h and t in entity pair (h, t) could emerge from across different sentences S_h and S_t in the document $\mathcal{D}$.

Additionally, the document graph consists of heterogeneous types of nodes and edges in comparison with the sentence-level graph that contains only entity-nodes and single edge types among them.

IV. PROPOSED MODEL

In this section, we introduce the overall framework of our DoreBer networks, as shown in Figure 2.

A. CONTEXT EMBEDDING MODULE

GloVe (Global Vectors) [58] is a pre-trained language model proposed by Pennington in 2014, aiming for obtaining vector representations for words. Since then, GloVe has been widely used in various NLP tasks. Training process of GloVE is performed on clustered global word-word co-occurrence matrix, which tabulates the frequencies of words co-occur with the one another in a given corpus. Population of the co-occurrence matrix requires a single pass through the whole corpus to calculate the statistics. For large corpora, the training process of GloVE can be computationally expensive, but it is just a one-time cost in preprocessing period.

In our approach, we use GloVe in embedding layer to provide vector representations for words. The parameter of GloVe is initialized by the pre-trained parameters provided by [58].

B. RECURRENT NEURAL NETWORK ENCODING MODULE

Recent studies have shown that using Recurrent Neural Networks as encoder to produce features has achieved great success in many NLP tasks. Specifically, our DoreBer model is based on GRU (Gated Recurrent Unit) [59] to perform recurrent neural network encoding.

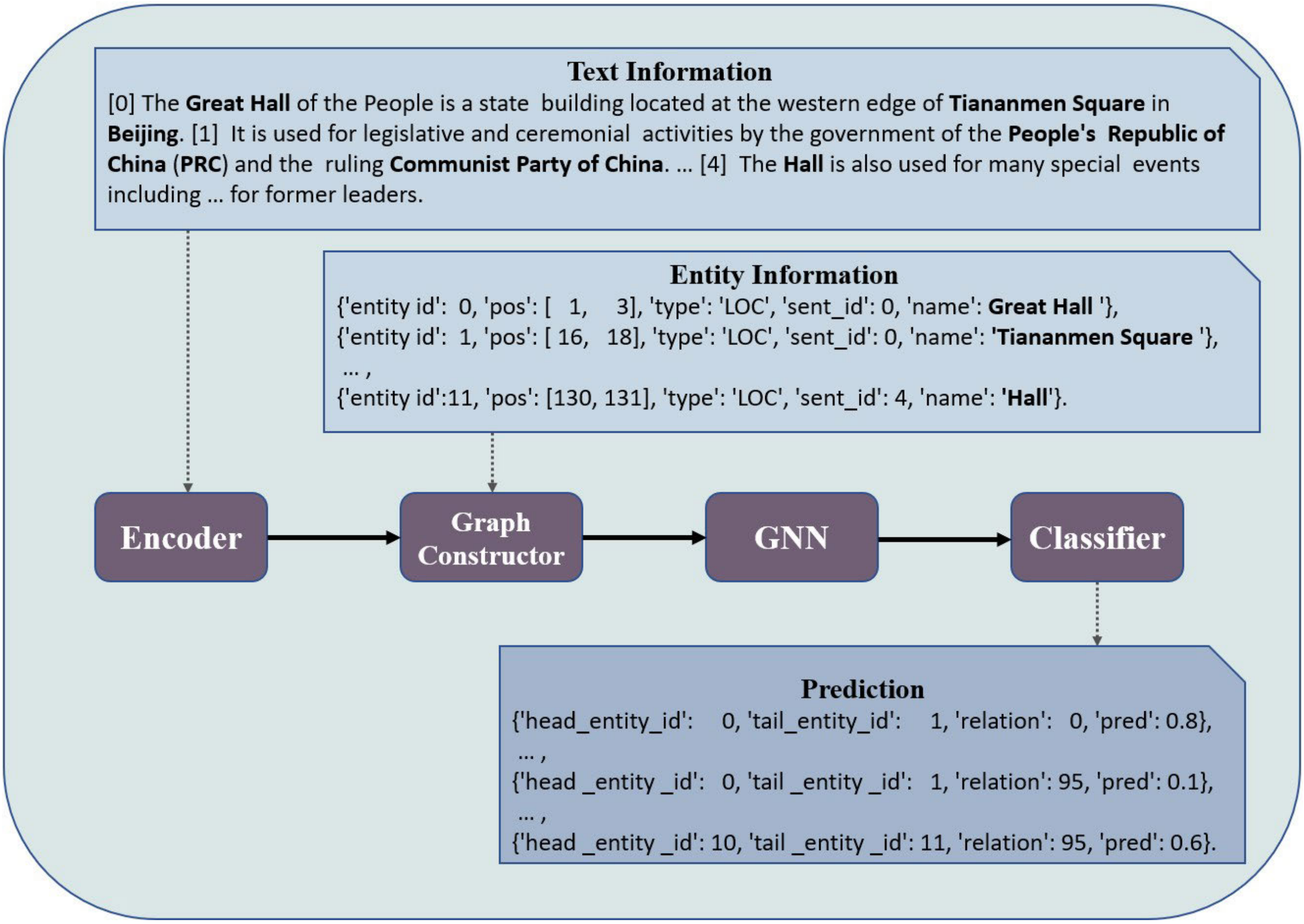

FIGURE 2. **Architecture of DoreBer. First, the context encoder process the text information to get contextualized embedding of each word. Then, a document-level graph is constructed by the graph constructor with diverse information from the encoder and the newly entered entity information. After applying GNN, presentations of each node in the document-level graph will be updated by using information aggregation scheme. Finally, the classifier module predicts multi-label relations based on the updated presentations.**

C. NODES EXTRACTION MODULE

In the nodes extraction module, we construct nodes for a document-level graph in three types following [24]: mention nodes, entity nodes and meta dependency path(MDP) nodes.

Mention nodes refer to different mentions of named entities in a document. And the representation of a mention node is formulated as the average of the representations of its span of tokens. **Entity nodes** refer to different mentions represented the same named entity. And the representation of an entity node is calculated as the average of its mentions' representation. To build a better document-level graph, information in dependency tree of a sentence has been introduced into. Previous approaches leverage all mention nodes in the dependency tree of a sentence [42] or introduce a sentence-level node by averaging representations of all mention nodes in a sentence. Following [24], we only leverage the tokens on the shortest dependency path for mention pairs in the sentence, which has been widely used in sentence-level relation extraction task as it is benefit to utilize relevant information effectively while masking irrelevant one [33], [60]. **Meta dependency nodes** refer to tokens on the shortest dependency path between mentions in the same sentence. And the representation of an meta dependency node is the same as its corresponding token.

D. GRAPH REASONING MODULE

Graph reasoner module has two sub-modules: graph structure induction sub-module and graph representation extraction sub-module. The graph structure induction sub-module is designed to learn a representation of a document-level graph. And the graph representation extraction sub-module is proposed to perform graph inference on the induced representation of a document-level graph, where representations of each node will be updated by using information aggregation scheme. Following [24], we stack N blocks of Graph reasoner module in order to iteratively update the representation of a document-level graph for better reasoning.

1) GRAPH STRUCTURE INDUCTION SUB-MODULE

Let $\boldsymbol{u}_i$ denote the contextual representation of the i-th node, where $\boldsymbol{u}_i \in \mathbb{R}^d$, we first calculate the pair-wise unnormalized attention score $\boldsymbol{s}_{ij}$ between the i-th and the j-th node with the node representations $\boldsymbol{u}_i$ and $\boldsymbol{u}_j$. The score $\boldsymbol{s}_{ij}$ is calculated by two feed-forward neural networks and a bilinear transformation:

$$\boldsymbol{s}_{ij} = (\tanh(\boldsymbol{W}_p\boldsymbol{u}_i))^T\boldsymbol{W}_b(\tanh\boldsymbol{W}_c\boldsymbol{u}_j)), \tag{1}$$

where $\boldsymbol{W}_p \in \mathbb{R}^{d\times d}$ and $\boldsymbol{W}_c \in \mathbb{R}^{d\times d}$ are weights for two feed-forward neural networks, d is the dimension of the node representations, and tanh is applied as the activation function. $\boldsymbol{W}_b \in \mathbb{R}^{d\times d}$ are the weights for the bilinear transformation. Next we compute the root score $\boldsymbol{s}_i^r$ which represents the unnormalized probability of the i-th node to be selected as the root node of the structure:

$$\boldsymbol{s}_i^r = \boldsymbol{W}_r\boldsymbol{u}_i \tag{2}$$

is the weight for the linear transformation. Following [28], we calculate the marginal probability of each dependency edge of the document-level graph. For a graph $\boldsymbol{G}$ with n nodes, we first assign non-negative weights $\boldsymbol{P} \in \mathbb{R}^{d\times d}$ to the edges of the graph:

$$\boldsymbol{P}_{ij} = \begin{cases} 0, & \text{if } i = j, \\ \exp(\boldsymbol{s}_{ij}), & \text{otherwise,} \end{cases} \tag{3}$$

where $\boldsymbol{P}_{ij}$ is the weight of the edge between the i-th and the j-th node. We then define the Laplacian matrix $\boldsymbol{L} \in \mathbb{R}^{n\times n}$ of $\boldsymbol{G}$ in Equation (6), and its variant $\hat{\boldsymbol{L}} \in \mathbb{R}^{n\times n}$ in Equation (7) for further computations [28]:

$$\boldsymbol{L}_{ij} = \begin{cases} \sum_{i'=1}^{n} \boldsymbol{P}_{i'j}, & \text{if } i = j, \\ -\boldsymbol{P}_{ij}, & \text{otherwise,} \end{cases} \tag{4}$$

$$\hat{\boldsymbol{L}}_{ij} = \begin{cases} \exp(\boldsymbol{s}_i^r), & \text{if } i = j, \\ \boldsymbol{L}_{ij}, & \text{if } i > j. \end{cases} \tag{5}$$

We use $\boldsymbol{A}_{ij}$ to denote the marginal probability of the dependency edge between the i-th and the j-th node. Then, $\boldsymbol{A}_{ij}$ can be derived based on Equation (8), where δ is the Kronecker delta [28]:

$$\boldsymbol{A}_{ij} = (1-\delta_{1,j})\boldsymbol{P}_{ij}[\hat{\boldsymbol{L}}^{-1}]_{ij} - (1-\delta_{i,1})\boldsymbol{P}_{ij}[\hat{\boldsymbol{L}}^{-1}]_{ji}. \tag{6}$$

Here, $\boldsymbol{A} \in \mathbb{R}^{d\times d}$ can be interpreted as a weighted adjacency matrix of the document-level entity graph. Finally, we can feed $\boldsymbol{A} \in \mathbb{R}^{d\times d}$ into the multi-hop reasoning module to update the representations of nodes in the latent structure.

2) GRAPH REPRESENTATION EXTRACTION SUB-MODULE

Graph neural networks (GNNs) [52] have been widely used in different task to extract representation(GCN [22], ChebNet [61],), due to their excellent ability to collect relevant signals based on an information aggregation scheme. Specifically, our DoreBer model is based on BernNet [23] to extract graph representation.

Formally, given a graph with n nodes, which can be represented with an $n \times n$ adjacency matrix $\boldsymbol{A}$ induced by the previous structure induction module:

$$\boldsymbol{Z} = \sum_{k=0}^{K} \theta_k \frac{1}{2^K}\binom{K}{k}(2\boldsymbol{I}-\boldsymbol{L})^{K-k}\boldsymbol{L}^k f(\boldsymbol{X}), \tag{7}$$

where $f(\boldsymbol{X})$ is a 2-layer MLP with 64 hidden units on the feature matrix $\boldsymbol{X}$.

E. ITERATIVE REFINEMENT MODULE

Though structured attention [25], [26] is capable to automatically induce structure, recent study show that the structure generated by structured attention is relatively shallow and may not be capable to capture the intricate dependencies for the document-level task [38], [62]. Following [24], we iteratively update the document-level graph based on the previously updated graph representations rather than inducing the latent structure only once, allowing the model to infer structure into a more sophisticated style that could provide more information than simple parent-child relations.

We stack N blocks of the graph reasoning module for iteratively inducing the document-level graph N times. Specifically, the structure induced at early iterations is relatively shallow, because the node representation was propagated mostly among the neighboring nodes. With the iteration of induction, the structure absorbs more information from non-local interactions, making the induction sub-module be capable to generate a more informative graph.

F. RELATION CLASSIFICATION MODULE

After N times of iterative refinement, we get $\boldsymbol{X}_{GNN}$, the representations of all the nodes extracted by GNN [52]. Now, we merge together representations of all the nodes extracted by RNN [63] and GNN [52], $\boldsymbol{X}_{RNN}$ and $\boldsymbol{X}_{GNN}$, through direct sum operation to get $\boldsymbol{X}$, the final representations of all the nodes extracted:

$$\boldsymbol{X} = \boldsymbol{X}_{RNN} \oplus \boldsymbol{X}_{GNN}, \tag{8}$$

where $\oplus$ is the direct sum operation. Next we calculate the representations of all the head and tail entity nodes, $\widetilde{\boldsymbol{e}}_{head}$ and $\widetilde{\boldsymbol{e}}_{tail}$, which are composed of weighted sum of all its mention representation(i.e. node representation):

$$\begin{aligned} \widetilde{\boldsymbol{e}}_{head} &= \boldsymbol{H} \cdot \boldsymbol{X}, \\ \widetilde{\boldsymbol{e}}_{tail} &= \boldsymbol{T} \cdot \boldsymbol{X}, \end{aligned} \tag{9}$$

where $\boldsymbol{H}$ and $\boldsymbol{T}$ denote the head and tail entity mapping matrices, which element denotes the weight of each node of each mention of each entity:

$$\boldsymbol{H}_{i,k} = \begin{cases} 1/(\boldsymbol{L}_{ent}^i \cdot \boldsymbol{L}_{men}^j), & \text{if } \textbf{node}^{i,k} \text{ belongs to } \\ & \textbf{mention}^j \text{ of } \textbf{entity}^i, \\ 0, & \text{otherwise,} \end{cases} \tag{10}$$

$$\boldsymbol{T}_{i,k} = \begin{cases} 1/(\boldsymbol{L}_{ent}^i \cdot \boldsymbol{L}_{men}^j), & \text{if } \textbf{node}^{i,k} \text{ belongs to } \\ & \textbf{mention}^j \text{ of } \textbf{entity}^i, \\ 0, & \text{otherwise,} \end{cases} \tag{11}$$

where L_{ent}^{i} and L_{men}^{j} represent the mention count of the i^{th} entity and the length of the j^{th} mention. Then, we calculate $\boldsymbol{e_{head}}$ and $\boldsymbol{e_{tail}}$, the final representations of head and tail entity nodes:

$$\begin{aligned} \boldsymbol{e_{head}} &= [\widetilde{\boldsymbol{e}}_{head}; \boldsymbol{E}(d_{ht})], \\ \boldsymbol{e_{tail}} &= [\widetilde{\boldsymbol{e}}_{tail}; \boldsymbol{E}(d_{th})], \end{aligned} \tag{12}$$

where $[\cdot;\cdot]$ denotes concatenation operation, d_{ht} and d_{th} are relative distances from head entity to tail entity and the inverse respectively, $\boldsymbol{E}$ is an embedding matrix of relative distance.

Following [21], we treat relation prediction between the given entity pair as a multi-label classification task. Specifically, we use a bilinear function to compute the probability for each predefined relation type r as:

$$P(r|\boldsymbol{e}_i, \boldsymbol{e}_j) = \sigma(\boldsymbol{e}_i^T \boldsymbol{W_r} \boldsymbol{e}_j + \boldsymbol{b}_r), \tag{13}$$

where $\boldsymbol{W_r} \in \mathbb{R}^{d\times k\times d}$ and $b_r \in \mathbb{R}^k$ are relation type dependent trainable parameters, with k being the number of predefined relation categories, σ is the *ReLU* [64] function.

V. EXPERIMENTS

In this part, the performance of proposed method is evaluated and compared with the previous sequence-based state-of-art document-level relation extraction models.

A. DATASET AND EVALUATION METRICS

In this paper, we select a popular benchmark dataset, DocRED [21], to assess our DoreBer model. DocRED is a large document-level dataset for tasks of relation extraction, named entity recognition and so on. DocRED made use of Wikidata, a large-scale knowledge base tightly integrated with Wikipedia. And the labeld data contains named entity recognition results, entity linking results and relation information, which could provide useful information in downstream processing.

Docred has two parts: human-annonated part and distantly supervised part. We evaluate DoreBer on the human-annonated part of DocRED dataset, which contains 5,053 documents, 1,002k words, 40,276 sentences, 132,375 entities, 96 pre-defined relations, 63,427 instances and 56,354 facts, as shown in Table 2. Following [21], we split the dataset into 3053/1000/1000 documents for traininng, development and test sets.

Following [24], we constructed 3 types of nodes for a document-level graph: mention nodes, entity nodes and meta dependency paths (MDP) nodes.

The coverage of the DocRED is comprehensive. The 96 frequent relation types are chosen from Wikidata, containing but not limited to logical relations, geographical relations and human-related relations. It is a need of reasoning for identifying most of relation instances in the DocRED, even logical reasoning, coreference reasoning, common-sense reasoning and pattern recognition.

TABLE 2. Statistics of data used for experiments.

Documents	5,053
Words	1,002K
Sentences	40,276
Entities	132,375
Relations	96
Instances	63,427
Facts	56,354

In DocRED dataset, each relation instance is correlated with 1.6 supporting sentences averagely, where 46.4% relation instances are correlated with more than one supporting instance. Additionally, there exist 40.7% relation facts which could **only** be identified through multiple sentences. Therefore, with a large amount of inter-sentential relation facts, DocRED dataset becomes appropriate for assessing document-level relation extraction approches.

B. EXPERIMENT SETTINGS

Following [24], we use spaCy[1] to get meta dependency paths of sentences in a document.

Following [21] and [29], we use the [58] embedding with BiLSTM as context encoder. And all hyper-parameters are adjusted based on the development set. We list part of important hyper-parameters in Table 3.

Evaluation Metrics: Following [21], two widely used metrics F_1 and Ign F_1 are used in our experiments. Ign F_1 denotes F_1 scores except relational facts shared by the training and dev/test sets. particularly, evaluation of our DoreBer model on test dataset was conducted through CodaLab.[2] The calculation formula for evaluation metrics are shown as blow:

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}, \tag{14}$$

$$\text{Ign } F_1 = 2 \cdot \frac{\text{precision}_{\text{ign}} \cdot \text{recall}_{\text{ign}}}{\text{precision}_{\text{ign}} + \text{recall}_{\text{ign}}}, \tag{15}$$

where $\text{precision}_{\text{ign}}$ and $\text{recall}_{\text{ign}}$ scores are precision and recall scores excluding those relational facts shared by the training and dev/test sets.

C. OVERALL EVALUATION RESULTS

We compare our proposed model with following two types of competitive models on the DocRED dataset, and show the main result in Table 4.

- **Sequence-based Models.** These models use different neural architectures to encode sentences in the document, including a CNN based model [15], an LSTM based model, a bidirectoinal LSTM (BiLSTM) based model [65] and an attention-based LSTM model (Context-Aware) [66]. These there models differ only at the encoder used for encoding the document.

[1]https://spacy.io
[2]https://codalab.lisn.upsaclay.fr/competitions/365

TABLE 3. Hyper-parameters of DoreBer.

Batch size	10
Learning rate	0.001
Optimizer	Adam
Hidden size	120
Induction block number	2
GCN dropout	0.3

- **Graph-based Models.** These models construct document-specific graphs for relation extraction. We adapt GCNN [42], EoG [43], GAT [57], AGGCN [67] to document-level RE scenario.

As shown in Table 4, our proposed model DoreBer achieves 52.31 Ign F_1 and 54.58 F_1 on the test set, which is the new state-of-the-art result for GloVe-based methods. Particularly, DoreBer consistently outperforms the sequence-based models in both Ign F_1 and F_1 by a large margin. For example, our DoreBer improves upon the best sequence-based model BiLSTM by 3.53 points in terms of Ign F_1 and 3.52 points in terms of F_1 on test set. This suggest that models which directly encode the entire document without leveraging document-specific graphs are unable to capture non-local dependencies in the document.

Under the same setting, DoreBer consistently outperforms graph-based models. DoreBer achieves 2.72 and 2.76 higher Ign F_1 and F_1 on test set, respectively, comparing with the best graph-based model EoG. This shows that previous graph methods may not be able to capture the complex interactions in a document.

D. ABLATION STUDY

1) MLP

Table 5 shows F_1 scores of DoreBer with and without MLP in graph representation extraction sub-module of graph reasoning module. The MLP has 2 layers, the first layer mapping features from 120 dims to 32 dims, and the second one mapping back to 120 dims. And both layers are equiped with Relu activation function, making mapping nonlinear to be capable to fit intricate functions. The MLP, which first conduct dimension reduction and then ascendance, is able to extract critical features. We observe that the MLP contributes to the main model, as the performance deteriorates with the MLP missing. Removal of MLP leads to a 15.29% drop in terms of F_1 score and 16.03% in terms of Ign F_1 score. Therefore, we obtain the conclusion that MLP plays a key role in extracting features in graph reasoning module.

2) PARAMETER COUNT

In this section, we will illustrate whether the larger parameter count of the Bernstein polynomial in the BernNet [23] layer brings improvement in the evaluation metrics, i.e. Ign F_1 and F_1.

TABLE 4. Main results on the development set and test set of DocRED dataset. Models with ♣ are adapted to DocRED based on their implementations.

	Dev		Test	
Model	Ign F_1	F_1	Ign F_1	F_1
CNN ♣	41.58	43.45	40.33	42.26
LSTM ♣	48.44	50.68	47.71	50.07
BiLSTM ♣	48.87	50.94	48.78	51.06
ContexAware ♣	48.94	51.09	48.40	50.70
GCNN ♣	46.22	51.52	49.59	51.62
EoG ♣	45.94	52.15	49.48	51.82
GAT ♣	45.17	51.44	47.36	49.51
AGGCN ♣	46.29	52.47	48.89	51.45
GloVe+LSR ♣	48.82	55.17	52.15	54.18
DoreBer	**49.24**	**55.49**	**52.31**	**54.58**

We test the following 6 experiments to confirm the assumption. Each experiment uses 10, 20, 40, 60, 80, and 100 parameters, respectively, in Bernstein polynomial of the BernNet layer.

The result are shown in Table 6. We can known that the rise of parameter amount in the BernNet layer, which leads a larger parameter number in total than the one in the previous GCN [22] layer, contributes to the main model. Additionally, the result demonstrates that the model performance monotonically improves, in general, with the increase of parameter amount in the BernNet layer. This empirically confirms our hypothesis that with an increase in amount of BernNet layers DoreBer is able to capture more informative features for the whole graph.

3) GNN LAYER

The effectiveness of the GNN [52] layer count has been verified by the experiment above. In this section, we test the different GNN layer counts to show whether it can brings improvement to our proposed model. We adopt the several layer settings which are 7, and 1. The result is shown in Table 7.

From Table 7, we can conclude that the rise in GNN layer amount can bring enormous benefits to DoreBer. What consistent with previous study [23] is that the BernNet is capable of alleviating over-smoothing problem. Over-smoothing problem means coefficients in GNN tend to be 0 with the amount of layer rising, leading a inefficiency in learning features. Hence, GCN could only use relative small amount of layers due to over-smoothing problem. However, BernNet could leverage a relative large amount of layer to extract features. Therefore, we obtain the conclusion that the higher GNN layer count can achieve better performance than single layer.

E. CASE STUDY

In this part, we perform a case study to further illustrate that our DoreBer model is capable to effectively capture the

TABLE 5. Ablation study of MLP of DoreBer on DocRED.

	Dev			Test	
Model	Ign F_1	F_1	AUC	Ign F_1	F_1
Full model	**49.24**	**55.49**	**55.63**	**52.31**	**54.58**
-MLP	34.00	39.47	34.04	33.99	26.31

TABLE 6. Ablation study of parameter count of DoreBer on DocRED.

	Dev			Test	
Model	Ign F_1	F_1	AUC	Ign F_1	F_1
Full model(K=100)	**49.24**	**55.49**	**55.63**	**52.31**	**54.58**
K=80	49.01	55.19	55.45	49.01	45.35
K=60	49.20	55.35	55.63	49.19	45.49
K=40	49.13	55.37	55.31	49.13	45.10
K=20	49.04	55.25	55.24	49.03	45.14
K=10	48.92	55.24	55.44	48.91	45.22

TABLE 7. Ablation study of GNN layer of DoreBer on DocRED.

	Dev			Test	
Model	Ign F_1	F_1	AUC	Ign F_1	F_1
Full model(layer=7)	**49.24**	**55.49**	**55.63**	**54.58**	**52.31**
layer=1	34.00	39.47	34.04	33.99	26.31

interaction among entity pairs and conduct document-level logical reasoning.

We take an example from development set of DocRED and visualize it in Figure 3. We use the entity *Loud Tour* and *United Kingdom* to demonstrate the reasoning process and our targets are to predict the relation type between <*Loud Tour, Rihanna*>, <*London, United Kingdom*>, and <*O2 Arena, United Kingdom*>. As shown in Figure 3, itLoud Tour and *Rihanna* appear in the 0^{th} sentence simultaneously, which is still a sentence-level RE task. And the relation can be correctly predicted as ''performer'', which is denoted by *P175*. For the second target, *London* emerges in the 4^{th} sentence, which is far away from *United Kingdom* in the 3^{th} sentence and relation prediction between them is up to the standard of document-level RE. Specifically, *Loud Tour* and *Rihanna* first interact with their local mentions respectively, and then be leveraged along with their relation predicted before as a bridge between *London* and *United Kingdom*. With inter- and intra-sentence information, the relation of <*London, United Kingdom*> is correctly predicted as ''country''. For the third target, *O2 Arena* arises in the 4^{st} sentence, which relation with <*United Kingdom*> is harder to predict due to the information is difficult to pass through one more bridge, <*London, United Kingdom*>. The result shows that the relation can be correctly predicted as ''country'', which indicates that DoreBer could capture the correlation across entity-pairs. The above results indicate that

Loud Tour

[0] The ***Loud Tour*** was the fourth overall and third world concert tour by Barbadian recording artist ***Rihanna***.

[3] The Loud Tour was a large commercial success, experiencing demand for an extension of shows in the ***United Kingdom*** due to popularity.

[4] In ***London***, Rihanna played a record breaking 10 dates at The ***O2 Arena***.

Subject: ***Loud Tour***
Object: ***Rihanna***
Relation: performer **Supporting Evidence:** 0

Subject: ***London***
Object: ***United Kingdom***
Relation: country **Supporting Evidence:** 3, 4

Subject: ***O2 Arena***
Object: ***United Kingdom***
Relation: country **Supporting Evidence:** 3, 4

FIGURE 3. Case study of an example from the development set of DocRED dataset.

DoreBer could effectively perform document-level RE and could be capable of multi-hop reasoning.

VI. CONCLUSION

In this paper, we introduce a novel model for better reasoning in the document-level RE task, which is referred to as DoreBer. Unlike previous approaches that rely on GCN, ChebNey or GPR-GNN, our DoreBer model leverage BernNet to aggregate graph feature. Online experimental results demonstrate that DoreBer achieves 2.72 and 2.76 higher Ign F_1 and F_1 on test set, respectively, comparing with the previous state-of-the-art. There exist several ways for future work. One possible direction is to utilize contrastive learning methods to enhance model learning ability.

REFERENCES

[1] B. D. Trisedya, G. Weikum, J. Qi, and R. Zhang, ''Neural relation extraction for knowledge base enrichment,'' in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, 2019, pp. 229–240.
[2] B. Min, R. Grishman, L. Wan, C. Wang, and D. Gondek, ''Distant supervision for relation extraction with an incomplete knowledge base,'' in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Human Lang. Technol.*, 2013, pp. 777–782.
[3] S. Jia, M. Li, and Y. Xiang, ''Chinese open relation extraction and knowledge base establishment,'' *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 17, no. 3, pp. 1–22, Sep. 2018.
[4] Y. Lin, Z. Liu, M. Sun, Y. Liu, and X. Zhu, ''Learning entity and relation embeddings for knowledge graph completion,'' in *Proc. AAAI Conf. Artif. Intell.*, 2015, vol. 29, no. 1, pp. 1–7.
[5] Y. Shen, N. Ding, H.-T. Zheng, Y. Li, and M. Yang, ''Modeling relation paths for knowledge graph completion,'' *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 11, pp. 3607–3617, Nov. 2021.
[6] B. Shi and T. Weninger, ''Open-world knowledge graph completion,'' in *Proc. AAAI Conf. Artif. Intell.*, 2018, vol. 32, no. 1, pp. 1–8.
[7] X. Hu, Z. Hong, C. Zhang, I. King, and P. Yu, ''Think rationally about what you see: Continuous rationale extraction for relation extraction,'' in *Proc. 46th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jul. 2023, pp. 2436–2440.
[8] J. Jiang, ''Information extraction from text,'' in *Mining Text Data*. New York, NY, USA: Springer, 2012, pp. 11–41.

[9] E. E. D. Loper, "Applying semantic relation extraction to information retrieval," Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, MA, USA, 2000.
[10] H. Elsahar, P. Vougiouklis, A. Remaci, C. Gravier, J. Hare, F. Laforest, and E. Simperl, "T-rex: A large scale alignment of natural language with knowledge base triples," in *Proc. 11th Int. Conf. Lang. Resour. Eval.*, 2018, pp. 1–5.
[11] S. Kim, M. Jeong, and G. G. Lee, "A local tree alignment approach to relation extraction of multiple arguments," *Inf. Process. Manag.*, vol. 47, no. 4, pp. 593–605, Jul. 2011.
[12] E. Beisswanger, "Exploiting relation extraction for ontology alignment," in *The Semantic Web—ISWC*. Shanghai, China: Springer, 2010, pp. 289–296.
[13] L. Bauer, Y. Wang, and M. Bansal, "Commonsense for generative multi-hop question answering tasks," *arXiv:1809.06309*, 2018.
[14] V. Gupta, M. Chinnakotla, and M. Shrivastava, "Transliteration better than translation? Answering code-mixed questions over a knowledge base," in *Proc. 3rd Workshop Comput. Approaches Linguistic Code-Switching*, 2018, pp. 39–50.
[15] D. Zeng, K. Liu, S. Lai, G. Zhou, and J. Zhao, "Relation classification via convolutional deep neural network," in *Proc. 25th Int. Conf. Comput. Linguistics, Tech. Papers*, 2014, pp. 2335–2344.
[16] K. Xu, Y. Feng, S. Huang, and D. Zhao, "Semantic relation classification via convolutional neural networks with simple negative sampling," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2015, pp. 1–5.
[17] N. Peng, H. Poon, C. Quirk, K. Toutanova, and W.-T. Yih, "Cross-sentence N-ary relation extraction with graph LSTMs," *Trans. Assoc. Comput. Linguistics*, vol. 5, pp. 101–115, Dec. 2017.
[18] C. Quirk and H. Poon, "Distant supervision for relation extraction beyond the sentence boundary," in *Proc. 15th Conf. Eur. Chapter Assoc. Comput. Linguistics*, 2017, pp. 1171–1182.
[19] P. Gupta, S. Rajaram, H. Schütze, and T. Runkler, "Neural relation extraction within and across sentence boundaries," in *Proc. AAAI*, vol. 33, 2019, pp. 6513–6520.
[20] L. Song, Y. Zhang, D. Gildea, M. Yu, Z. Wang, and J. Su, "Leveraging dependency forest for neural medical relation extraction," in *Proc. Conf. Empirical Methods Natural Lang. Process., 9th Int. Joint Conf. Natural Lang. Process. (EMNLP-IJCNLP)*, 2019, pp. 208–218.
[21] Y. Yao, D. Ye, P. Li, X. Han, Y. Lin, Z. Liu, Z. Liu, L. Huang, J. Zhou, and M. Sun, "DocRED: A large-scale document-level relation extraction dataset," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, 2019, pp. 764–777.
[22] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*.
[23] M. He, Z. Wei, and H. Xu, "BernNet: Learning arbitrary graph spectral filters via Bernstein approximation," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 14239–14251.
[24] G. Nan, Z. Guo, I. Sekulic, and W. Lu, "Reasoning with latent structure refinement for document-level relation extraction," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 1546–1557.
[25] Y. Liu and M. Lapata, "Learning structured text representations," *Trans. Assoc. Comput. Linguistics*, vol. 6, pp. 63–75, Dec. 2018.
[26] Y. Kim, C. Denton, and L. H. A. M. Rush, "Structured attention networks," 2017, *arXiv:1702.00887*.
[27] W. T. Tutte, "Graph theory," in *Encyclopedia of Mathematics and Its Applications*. CA, USA: Addison-Wesley, 1984.
[28] T. Koo, A. Globerson, X. C. Pérez, and M. Collins, "Structured prediction models via the matrix-tree theorem," in *Proc. Joint Conf. Empirical Methods Natural Lang. Process. Comput. Natural Lang. Learn. (EMNLP-CoNLL)*, 2007, pp. 141–150.
[29] L. Wang, Z. Cao, G. de Melo, and Z. Liu, "Relation classification via multi-level attention CNNs," in *Proc. 54th Annu. Meeting Assoc. Comput. Linguistics*, 2016, pp. 1298–1307.
[30] P. Zhou, W. Shi, J. Tian, Z. Qi, B. Li, H. Hao, and B. Xu, "Attention-based bidirectional long short-term memory networks for relation classification," in *Proc. 54th Annu. Meeting Assoc. Comput. Linguistics*, 2016, pp. 207–212.
[31] Y. Zhang, V. Zhong, D. Chen, G. Angeli, and C. D. Manning, "Position-aware attention and supervised data improve slot filling," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2017, pp. 1–16.
[32] Z. Guo, G. Nan, W. Lu, and S. B. Cohen, "Learning latent forests for medical relation extraction," in *Proc. 29th Int. Joint Conf. Artif. Intell.*, Jul. 2020, pp. 3651–3657.
[33] K. Xu, Y. Feng, S. Huang, and D. Zhao, "Semantic relation classification via convolutional neural networks with simple negative sampling," 2015, *arXiv:1506.07650*.
[34] Y. Liu, F. Wei, S. Li, H. Ji, M. Zhou, and H. Wang, "A dependency-based neural network for relation classification," in *Proc. 53rd Annu. Meeting Assoc. Comput. Linguistics, 7th Int. Joint Conf. Natural Lang. Process.*, 2015, pp. 285–290.
[35] M. Miwa and M. Bansal, "End-to-end relation extraction using LSTMs on sequences and tree structures," 2016, *arXiv:1601.00770*.
[36] Y. Zhang, P. Qi, and C. D. Manning, "Graph convolution over pruned dependency trees improves relation extraction," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2018, pp. 2205–2215.
[37] P. Verga, E. Strubell, and A. McCallum, "Simultaneously self-attending to all mentions for full-abstract biological relation extraction," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics*, 2018, pp. 872–884.
[38] Y. Liu, I. Titov, and M. Lapata, "Single document summarization as tree induction," in *Proc. Conf. North*, 2019, pp. 1745–1755.
[39] W. Lei, X. Wang, M. Liu, I. Ilievski, X. He, and M.-Y. Kan, "SWIM: A simple word interaction model for implicit discourse relation recognition," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, Aug. 2017, pp. 4026–4032.
[40] W. Lei, Y. Xiang, Y. Wang, Q. Zhong, M. Liu, and M.-Y. Kan, "Linguistic properties matter for implicit discourse relation recognition: Combining semantic interaction, topic continuity and attribution," in *Proc. AAAI Conf. Artif. Intell.*, 2018, vol. 32, no. 1, pp. 1–8.
[41] R. Jia, C. Wong, and H. Poon, "Document-level N-ary relation extraction with multiscale representation learning," in *Proc. Conf. North*, 2019, pp. 3693–3704.
[42] S. K. Sahu, F. Christopoulou, M. Miwa, and S. Ananiadou, "Inter-sentence relation extraction with document-level graph convolutional neural network," 2019, *arXiv:1906.04684*.
[43] F. Christopoulou, M. Miwa, and S. Ananiadou, "Connecting the dots: Document-level neural relation extraction with edge-oriented graphs," in *Proc. Conf. Empirical Methods Natural Lang. Process., 9th Int. Joint Conf. Natural Lang. Process. (EMNLP-IJCNLP)*, 2019, pp. 4925–4936.
[44] Y. Ma, A. Wang, and N. Okazaki, "DREEAM: Guiding attention with evidence for improving document-level relation extraction," 2023, *arXiv:2302.08675*.
[45] J. Giorgi, G. D. Bader, and B. Wang, "A sequence-to-sequence approach for document-level relation extraction," 2022, *arXiv:2204.01098*.
[46] J. Yu, D. Yang, and S. Tian, "Relation-specific attentions over entity mentions for enhanced document-level relation extraction," 2022, *arXiv:2205.14393*.
[47] S. Khan, H. U. Khan, S. Nazir, B. Albahooth, and M. Arif, "Users sentiment analysis using artificial intelligence-based FinTech data fusion in financial organizations," in *Mobile Networks and Applications*. Dordrecht, The Netherlands: Springer, Sep. 2023, pp. 1–12.
[48] Q. Tan, R. He, L. Bing, and H. T. Ng, "Document-level relation extraction with adaptive focal loss and knowledge distillation," 2022, *arXiv:2203.10900*.
[49] Y. Zhou and W. S. Lee, "None class ranking loss for document-level relation extraction," 2022, *arXiv:2205.00476*.
[50] W. Xu, K. Chen, L. Mou, and T. Zhao, "Document-level relation extraction with sentences importance estimation and focusing," 2022, *arXiv:2204.12679*.
[51] Y. Wei and Q. Li, "SagDRE: Sequence-aware graph-based document-level relation extraction with adaptive margin loss," in *Proc. 28th ACM SIGKDD Conf. Knowl. Discovery Data Mining*, Aug. 2022, pp. 2000–2008.
[52] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE Trans. Neural Netw.*, vol. 20, no. 1, pp. 61–80, Jan. 2008.
[53] X. Peng, C. Zhang, and K. Xu, "Document-level relation extraction via subgraph reasoning," in *Proc. 31st Int. Joint Conf. Artif. Intell.*, Jul. 2022, pp. 4331–4337.
[54] R. Zhang, Y. Li, and L. Zou, "A novel table-to-graph generation approach for document-level joint entity and relation extraction," in *Proc. 61st Annu. Meeting Assoc. Comput. Linguistics*, 2023, pp. 10853–10865.
[55] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.
[56] S. Zeng, R. Xu, B. Chang, and L. Li, "Double graph based reasoning for document-level relation extraction," 2020, *arXiv:2009.13752*.

[57] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *Stat*, vol. 1050, no. 20, p. 48550, 2017.

[58] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2014, pp. 1532–1543.

[59] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," in *Proc. NIPS*, Dec. 2014, pp. 1–9.

[60] R. C. Bunescu and R. J. Mooney, "A shortest path dependency kernel for relation extraction," in *Proc. Conf. Human Lang. Technol. Empirical Methods Natural Lang. Process.*, 2005, pp. 724–731.

[61] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 1–9.

[62] E. Ferracane, G. Durrett, J. J. Li, and K. Erk, "Evaluating discourse in structured text representations," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, 2019, pp. 646–653.

[63] J. L. Elman, "Finding structure in time," *Cognit. Sci.*, vol. 14, no. 2, pp. 179–211, Mar. 1990.

[64] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 807–814.

[65] R. Cai, X. Zhang, and H. Wang, "Bidirectional recurrent convolutional neural network for relation classification," in *Proc. 54th Annu. Meeting Assoc. Comput. Linguistics*, 2016, pp. 756–765.

[66] D. Sorokin and I. Gurevych, "Context-aware representations for knowledge base relation extraction," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2017, pp. 1784–1789.

[67] Z. Guo, Y. Zhang, and W. Lu, "Attention guided graph convolutional networks for relation extraction," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, 2019, pp. 241–251.

BOYA YUAN was born in Hunan, China, in 1998. He received the B.S. degree in statistics from the Hunan University of Science and Technology, in 2020. He is currently pursuing the M.S. degree in statistics major with the College of Science, North China University of Technology. His current research interests include relation extraction, graph neural networks, transfer learning, and deep learning.

LIWEN XU was born in Anhui, China, in 1977. He received the B.S. degree in mathematics from the Changsha Institute of Electric Power, Changsha, in 2000, the M.S. degree in applied mathematics from Hunan University, Changsha, in 2003, and the Ph.D. degree in probability and mathematical statistics from the Beijing University of Technology, Beijing, in 2006. From 2006 to 2008, he was a Postdoctoral Researcher with the Department of Mathematical Science, Tsinghua University. Since 2008, he has been an Assistant Professor with the Department of Statistics, North China University of Technology, where he was a Full Professor in statistics, in 2015. He is the author of three books and more than 50 articles. His research interests include decentralized learning over multitask networks, deep learning, subsampling, and smoothing spline.