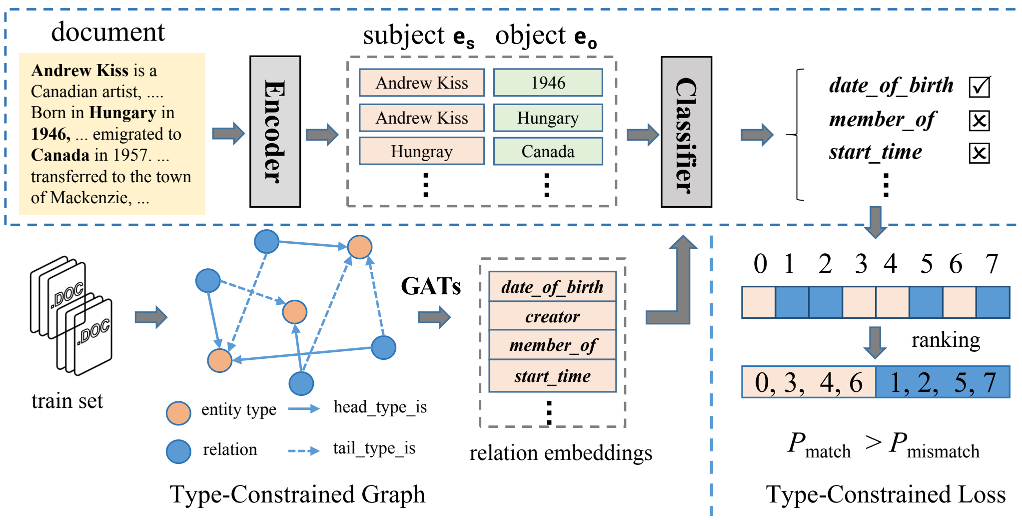
Graphical Abstract

Document-level Relation Extraction with Entity Type Constraints

Ridong Han, Tao Peng, BeiBei Zhu, Haijia Bi, Jiayu Han, Xinzheng Xu, Lu Liu



Highlights

Document-level Relation Extraction with Entity Type Constraints

Ridong Han, Tao Peng, BeiBei Zhu, Haijia Bi, Jiayu Han, Xinzheng Xu, Lu Liu

- Using relation correlations to solve both long-tail and multi-label problems.
- Capturing relation correlations with entity type constraints from the global and local perspectives.
- Type-constrained graph is constructed in the global perspective.
- Type-constrained loss is designed in the local perspective.
- Experiments confirm the effectiveness of proposed model in solving long-tail problem and multi-label problem.

Document-level Relation Extraction with Entity Type Constraints

Ridong Han[a,b,d], Tao Peng[b,c,d,*], BeiBei Zhu[e], Haijia Bi[b,d], Jiayu Han[f], Xinzheng Xu[a], Lu Liu[b,c,d,*]

[a]*School of Computer Science and Technology, China University of Mining and Technology, Xuzhou, 221116, Jiangsu, China*
[b]*College of Computer Science and Technology, Jilin University, Changchun, 130012, Jilin, China*
[c]*College of Software, Jilin University, Changchun, 130012, Jilin, China*
[d]*Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun, 130012, Jilin, China*
[e]*College of Computer and Artificial Intelligence, Liaoning Normal University, Dalian, 116081, Liaoning, China*
[f]*Department of Linguistics, University of Washington, Seattle, 98195, WA, United States*

Abstract

Long-tail problem and multi-label problem are two commonly encountered challenges in document-level relation extraction task. Current efforts are concerned with enhancing the representations of entity pairs through Transformer or document graphs, which cannot tackle the above challenges well. Relation correlations are a potential solution, which allows head relations to assist in the training of tail ones by transferring correlation knowledge between them, and can measure the semantic distance between relations to assist the classifier in assigning multiple semantically similar relations to

*Corresponding authors.

Email addresses: hanrd@cumt.edu.cn (Ridong Han), tpeng@jlu.edu.cn (Tao Peng), zhubb24@lnnu.edu.cn (BeiBei Zhu), bihj21@mails.jlu.edu.cn (Haijia Bi), jyhan126@uw.edu (Jiayu Han), xxzheng@cumt.edu.cn (Xinzheng Xu), liulu@jlu.edu.cn (Lu Liu)

multi-label instances. This paper proposes to learn relation correlations from both global and local views using entity type constraints, which means that the subject-object entity types limit the scope of possible relations. Globally, we statistically construct the Type-constrained Graph between entity types and relations, which formulates all possible subject/object types for each relation. Different relations are connected by common entity types, reflecting the desired correlations. Locally, given an entity pair, the classification probability of relations matching its entity types should be greater than those unmatched. Therefore, the Type-constrained Loss is proposed to make the matched relations have greater probabilities. Detailed experiments are conducted on DocRED and DWIE datasets, and our model significantly outperforms baselines under long-tail and multi-label setups.

Keywords: Relation Extraction, Document-level, Entity Type Constraints, Type-Constrained Graph, Type-Constrained Loss

---

1. Introduction

1.1. Background and Limitations

Relation extraction (RE) task, as a crucial stage of knowledge graph construction, has been extensively studied recently. It targets at identifying entities scattered in plain text and determining relationships between each pair of entities. A great deal of earlier studies focus on solving the simplest single-sentence scenario [1, 2, 3, 4, 5, 6], i.e., sentence-level relation extraction (SentRE), which determine the relationship of two given entities scattered in a sentence. Recently, some studies show that extensive relational facts do not exist within a single sentence, but are conveyed by several sentences within a document at the same time. Using the DocRED dataset [7] as an example, at least 40% triplets are represented by two entities scattered in multiple sentences. Therefore, sentence-level relation extraction is extended beyond sentence boundaries, i.e., document-level relation extraction (DocRE) that simultaneously determines the semantic relationships for all pairs of entities contained in a given document. The DocRE task has two seriously performance-impairing challenges, as follows:

- **Long-tail problem**: The amount of training instances[1] varies dramatically across pre-defined relation categories, which complies with a long-tailed distribution. Some relations are insufficiently trained and underfitted due to lack of training triplets, causing poor performance.
- **Multi-label problem**: According to the given context, Some pairs of entities simultaneously convey several target relation categories. These categories share some degree of semantic overlap, i.e., the semantic distance between them is closer than other relations. This requires the classifier to delineate the classification boundaries among relations more delicately.

Take the most commonly used DocRED dataset as an example, about 60 of the 96 relation categories have fewer than 200 triplets in the train set, which can be called long-tail categories. Such high ratio indicates the severe impact of long-tail problem. Additionally, in the train set, about 2500 entity pairs have multiple relation labels, and some of them even express four relations

[1] In this paper, an instance corresponds to an entity pair.

simultaneously. Multi-label entity pairs make up at least 7% of the dataset, which should not be ignored. However, existing efforts primarily concentrate on enhancing the contextual features of entity pairs through Transformer architecture [8, 9] or document graphs [10, 11, 12], which hardly solve the above challenges.

1.1.1. In the Background of Large Language Models

Given the dominance of large language models (LLMs) in the field of natural language processing, we explore the performance of ChatGPT [13] supported by GPT-3.5 and GPT-4 [14] on the DocRE task, and the results are displayed in Table 3. Specifically, `gpt-3.5-turbo` and `gpt-4-0125-preview` are used. From Table 3, it is clear that there is a huge gap between ChatGPT and existing baselines on the DocRE task, which is consistent with the conclusions of Li et al. [15] and Han et al. [16]. To investigate the capability of ChatGPT in solving the above two challenges, we launch experiments and present results in Table 4 and Table 5. It can be seen that ChatGPT is still far from existing baselines and cannot handle the above challenges well, regardless of GPT-3.5 or GPT-4. To sum up, despite the powerful capabilities of LLMs such as ChatGPT, they are still not capable of DocRE task, especially in solving the long-tail problem and multi-label problem. This paper aims to simultaneously solve the above two challenging problems by entity type constraints.

1.2. Motivation

Relation correlations are a potential solution to the above challenges, as follows:

(1) For long-tail problem, long-tail relations are usually correlated with some other relations, which may not be long-tailed. In other words, tail data-scarce relations can be related to some head data-rich ones. During training, by the relation correlations, the model can convey correlation knowledge from head categories to tail ones to facilitate the training of long-tailed categories, relieving the undertraining phenomenon.

(2) For multi-label problem, two entities convey multiple semantic relationships within the same context, which indicates that the semantic distance between these relations is much closer than other relations. Relation correlations provide a measure of the semantic distance across target categories and facilitate models to discriminate similar categories for multi-label entity pairs.

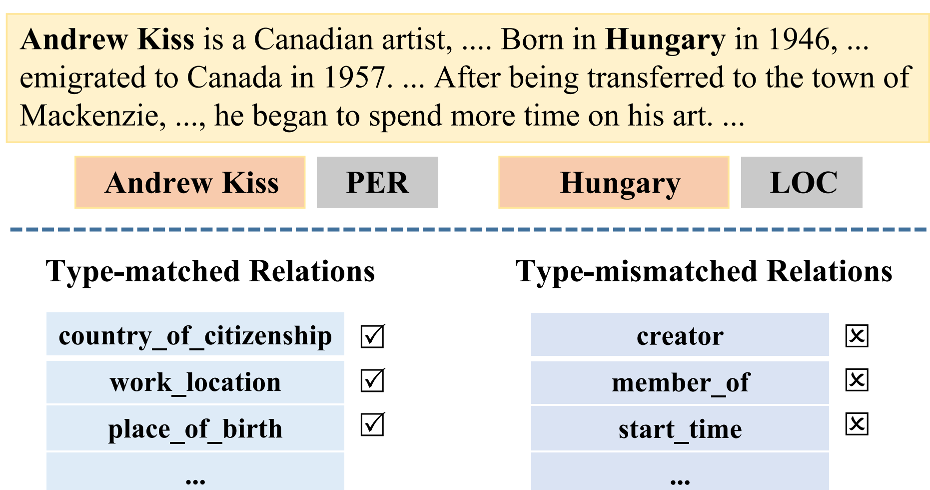


Figure 1: Given a biography, to determine the relationship between "Andrew Kiss" and "Hungary", all relations can be divided into type-matched set and type-mismatched set based on their type "PER" and "LOC".

Han et al. [17] and Huang et al. [18] employ the relation co-occurrence phenomenon to capture relation correlations, which is not delicate and may introduce noisy correlations, because the co-occurrence does not imply that the correlations necessarily exist. Different from them, we utilize the often-overlooked Entity Type Constraints (ETC) to model the correlations. Entity type constraints mean that the subject/object entity types limit the scope of possibly expressed relations, in other words, the types of subject and object entities allowed by a relation category are fixed. Therefore, entity type-constrained correlations exist between relations with the same subject-object type, which is more accurate. For example, in Figure 1, when recognizing the relationship between "**Andrew Kiss**" and "**Hungary**", the relations matching their type "PER" and "LOC" are more likely than those mismatching, i.e., "country_of_citizenship" and "place_of_birth" have higher probabilities to be expressed, while "creator" and "member_of" are impossible to express. These entity type constraints imply that there are type-constrained correlations among type-matched relations.

1.3. Research Objectives

Our primary research objective is to address both of the above challenges by modelling relation correlations through entity-type constraints. Our specific research objectives are summarized as:

(1) Modeling relation correlations using entity type constraints with the help of graph structure, and obtaining all relation embeddings.

(2) Constructing extra features for each entity pair to be categorized, based on the above embeddings.

(3) Using entity type constraints to constrain the classification probabilities, making the classifier focus more on the relation categories that matches the corresponding entity types.

Specifically, we utilize entity type constraints from both global and local perspectives. From the global view, we perform statistics on the train set, and construct the Type-Constrained Graph (TCG). The graph contains two kinds of nodes (including *entity types* and *relations*) and two types of edges (including *subject_type_is* and *object_type_is*), which formulates all possible subject/object types for each relation. Different relations are connected by common entity types, and relation correlations exist between relations with the same subject-object type. Then, the multi-head Graph Attention Networks (GATs) [19] is used to encode on this graph to obtain all relation embeddings, which are exploited to construct extra feature representations for each entity pair to be categorized. From the local view, unlike taking statistics on the entire train set, we consider any given entity pair (i.e., entity-pair level). We argue that *given an entity pair, the classification probability of relations matching its entity types should be greater than those unmatched*. In other words, the classifier should focus more on the relations matching the given entity types. Therefore, we define the ranking-based Type-Constrained Loss (TCL) to make the matched relations have greater probabilities, which is similar to Oksuz et al. [20]. Since our contributions consist of **T**ype-**C**onstrained **G**raph and Type-Constrained **L**oss, our proposed model is denoted by **DocRE-TCGL**.

We select two commonly used datasets to conduct experiments, including DocRED [7] and DWIE [21]. The results reveal that proposed DocRE-TCGL obtains consistent performance gains, and significantly outperforms all typical or up-to-date baselines under long-tail and multi-label setups, by up to 6.26% and 4.91%, respectively.

1.4. Contributions

To sum up, our main contributions include the following points:

- It is the first time that entity type constraints are used to capture relation correlations to solve both long-tail and multi-label problems in DocRE task, as far as we know.
- From the global view, we statistically construct the Type-Constrained Graph (TCG) to formulate all subject/object types for each relation, which yields all relation embeddings used to construct the additional features.
- From the local view, we design the Type-Constrained Loss (TCL), which makes the classifier focus more on relation categories that match the given entity types, with higher probabilities.
- Experiments on two benchmarks reveal that DocRE-TCGL model dramatically exceeds competitive baselines under long-tail and multi-label setups. The code is available on the Github site[2].

In the following narrative, Section 2 introduces related studies and the research scope, Section 3 presents basic definitions and the proposed DocRE-TCGL approach, Section 4 introduces experimental details and exhibits the

[2]`https://github.com/RidongHan/DocRE-TCGL`

performance, Section 5 summarizes our contributions and draws final conclusions.

2. Related Studies

2.1. Relation Extraction

Earlier researches center around the simplest sentence-level scenario, i.e., sentence-level relation extraction (SentRE). These approaches are still sequence-based models, which are built mainly based on Convolutional Neural Networks (CNNs) [22, 2, 23, 24], Recurrent Neural Networks (RNNs) [25, 26], Graph Neural Networks (GNNs) [27] or attention mechanism [2, 28, 29, 4, 30]. The above models still mainly concentrate on local information [31], including entity position, entity distance, etc., which are not sufficient for more complex document-level scenario.

Document-level relation extraction necessitates the cross-sentence long-distance dependencies and reasoning. In other words, to recognize the relationship for a given entity pair, the DocRE systems need to take into account all relevant information scattered throughout the document [32, 33, 34]. Currently, DocRE models can be broadly categorized into three types, i.e., sequence-based models, graph-based models and Transformer-based models. The sequence-based models [7] directly encode the entire document using traditional CNNs [35] and RNNs [36], which is the same as SentRE task and has worse performance. The graph-based models are much more complex [37, 10, 38, 39, 33, 40, 12, 34], which require manual construction of document graphs and employ graph neural networks [19] to integrate the information of entire document for classification, attaining higher performance. The

Transformer-based models directly utilize the pretrained Transformer-based language models to capture global dependencies throughout entire document [41, 8, 42, 43, 18, 17, 9], which do not rely on hand-crafted rules and receive lots of attention.

To alleviate the above two challenges in Section 1.1, several efforts have been made to design different training objectives. For instance, Tan et al. [44] propose the focal loss function in order to assign greater weights to long-tail categories, mitigating the under-training of tail categories, while Zhou et al. [41] and Zhou et al. [45] extend the binary cross-entropy loss into the adaptive threshold loss and none-class ranking loss, which allows the classifier to assign multiple labels to multi-label entity pairs.

2.2. Relation Correlations

None of the above approaches tackle both challenges in Section 1.1 simultaneously. Analogy with the commonly used correlations between labels [46, 47], relation correlations, also called "Relation of relations", and are first defined by Jin et al. [48]. Fu et al. [49] employ the relatedness among category prototypes to enhance the training procedure with instances from other datasets. Han et al. [24] and Peng et al. [4] solve the distantly supervised relation extraction by utilize the available hierarchical structure of relations to model the correlations. While such hierarchical structure does not exist on the DocRE datasets, modelling relation correlations is much more challenging. Han et al. [17] and Huang et al. [18] exploit the co-occurrence phenomenon of relations within a document to capture the co-occurrence correlations between different relations. This way is intuitive, but it tends to introduce noisy correlations, because the co-occurrence does not imply that the correlations necessarily exist.

2.3. Entity Type in Relation Extraction

Entity types are one of the classic features in relation extraction task [50], the most common usage is to directly utilize entity type embeddings to construct additional features [51, 7, 52, 53, 54], through concatenation operation, attention mechanism, etc. There are also attempts to extract entities and relations simultaneously by mapping entity types and relational categories into the same space [55, 56], or to make pre-trained language models sensitive to entity types using entity marker technique [57, 58]. Bai et al. [53] involve the concept of entity type constraints, but still utilize attention mechanism to fuse entity type embeddings with word embeddings. Different from the above approaches of enhancing feature representations by entity types, this paper exploits entity type constraints to capture the correlations among different relation categories.

2.4. Differences with existing researches

Our DocRE-TCGL model belongs to the category of Transformer-based methods, and the main differences with other studies are as follows:

- Unlike existing studies that neglect two challenges or address one of the challenges, this paper addresses both challenges simultaneously with the help of relation correlations.
- Unlike modeling relation correlations by the taxonomic structure or co-occurrence phenomenon among relations, this paper captures relation correlations by means of entity type constraints, as described in Section 1.2.

- Unlike the direct use of entity type embeddings, this paper focuses on the constraints of entity types on relation categories, reflecting the type-constrained correlations among relations.

3. Our Proposed Methodology

In this section, we present the detailed definition of DocRE task, and detail the proposed DocRE-TCGL that utilizes entity type constraints to mine relation correlations in order to solve both long-tail problem and multi-label problem simultaneously.

3.1. Task Formulation

Consider a document d containing N_t tokens, denoted as $\{w_i\}_{i=1}^{N_t}$, which includes N_e named entities represented by $\{e_i\}_{i=1}^{N_e}$. Each of these entities e_i appears N_m^i times in document d, and each occurrence corresponds to an entity mention, indicated as $\{m_j\}_{j=1}^{N_m^i}$. Unlike sentence-level relation extraction task that focuses on only one entity pair, the objective of DocRE task is to assign at least one relational label from the set $\mathcal{R} \cup \texttt{NA}$ to every pair of distinct entities $(e_s, e_o)_{s,o=1,\ldots,N_e;s\neq o}$. Here, $\mathcal{R}$ is the predefined set of relation categories of interest, while "NA" signifies the absence of any relationship between two specific entities. In essence, DocRE is a multi-label classification task on multiple entity pairs.

3.2. Overview

As illustrated in Figure 2, our proposed DocRE-TCGL includes three parts: base model, Type-Constrained Graph (TCG) and Type-Constrained Loss (TCL). Specifically, (1) the base model encodes the whole document and

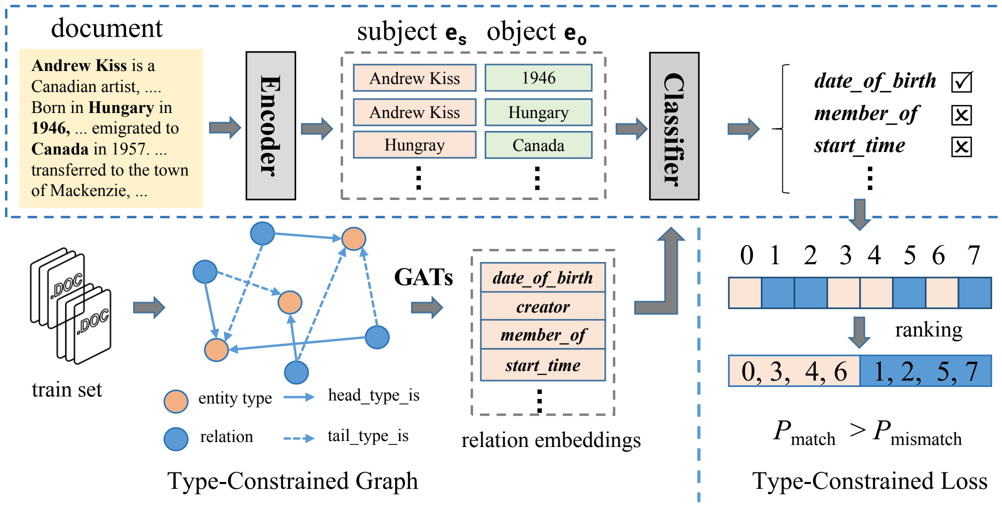


Figure 2: The architecture of DocRE-TCGL, which is comprised of three parts: the base model, Type-Constrained Graph and Type-Constrained Loss. The base model is circled by the blue dashed line, which encodes the document and categorizes every entity pair.

categories every entity pair within it, which is built on the existing models [59, 41] and is detailed in Section 3.3; (2) the Type-Constrained Graph aims to capture relation correlations from the global perspective, which formulates all possible subject/object entity types for each relation category and is encoded by the graph attention networks (GATs) to obtain all correlation-aware relation embeddings. These embeddings are then used to generate additional features for every pair of entities in order to guide the classifier in utilizing the correlation knowledge; (3) the Type-Constrained Loss intends to model relation correlations from the local perspective. For each entity pair, it makes the classification probabilities of relation categories matching its entity type greater than those mismatching.

3.3. Base Model

In theory, our DocRE-TCGL model does not depend on the structure of base model, and can utilize any existing model as our base model. To facilitate subsequent experimental comparisons in Section 4, we construct our base model (**DocRE-Base**) based on existing systems [59, 41].

Specifically, given a document $d = \{w_i\}_{i=1}^{N_t}$, it is encoded by a widely-used language model to generate all hidden embeddings of tokens,

$$H, A = \texttt{PLMs}([w_1, w_2, ..., w_{N_t}]) \tag{1}$$

where $\texttt{PLMs}(\cdot)$ denotes any pre-trained language models, $H = [h_1, h_2, ..., h_{N_t}] \in \mathbb{R}^{n \times d_h}$, d_h is the embedding dimension, and A is the multi-head attention weights from the last Transformer layer.

In the above process, the entity marker technique is employed to track the beginning and ending positions of all mentions with a special symbol "*", which has been confirmed to be very useful by several researches [60, 61]. Then, for each entity e_i with N_m^i mentions, the hidden representation of "*" in the prefix of entity mention m_j^i serves as its corresponding representation $h_{m_j^i}$. Entity embedding representation h_{e_i} can be obtained through log-sum-exp pooling operation on its all mentions, defined as,

$$h_{e_i} = \log \sum_{j=1}^{N_m^i} \exp\left(h_{m_j^i}\right) \tag{2}$$

where "log" and "exp" represent logarithmic and exponential operations respectively.

Following Zhou et al. [41], to capture context information for each entity pair (e_s, e_o), the above attention matrix A in Eq. 1 is utilized to aggregate all context information and obtain the contextual feature $c_{(s,o)} \in \mathbb{R}^{d_h}$. Specifically, for e_i, the attention weights of its mentions calculate the mean values as its weights $A_{e_i} \in \mathbb{R}^{N_h \times N_t}$. Here N_h denotes the number of attention heads. $c_{(s,o)}$ can be calculated as,

$$c_{(s,o)} = H^{\mathtt{T}} \cdot \mathtt{Norm}(\sum_{k=1}^{N_h} A_s^k \circ A_o^k) \tag{3}$$

where $\mathtt{Norm}(\cdot)$ denotes the normalization operation. Since all attention scores are always positive, the summation normalization is used here, defined as $\mathtt{Norm}(\vec{x}) = \vec{x}/\mathtt{sum}(\vec{x})$.

Finally, the grouped classifier is employed to complete the classification step for each entity pair [41]. Entity embeddings h_{e_s} and h_{e_o} are first enhanced with contextual representation $c_{(s,o)}$, respectively. Then, the resulting representations are input into the classifier.

$$[f_s^1; f_s^2; ...; f_s^k] = f_s = \tanh(W_s \cdot h_{e_s} + W_{c_1} \cdot c_{(s,o)}) \tag{4}$$

$$[f_o^1; f_o^2; ...; f_o^k] = f_o = \tanh(W_o \cdot h_{e_o} + W_{c_2} \cdot c_{(s,o)}) \tag{5}$$

$$P(r|e_s, e_o) = \sigma(\sum_{i=1}^{k} f_s^{i\mathtt{T}} \cdot W_r^i \cdot f_o^i + b_r) \tag{6}$$

where ";" is the concatenation operation between two tensors, k is the number of groups, W_s, W_o, W_{c_1}, W_{c_2} and $\{W_r^i\}_{i=1}^k$ are learnable parameters involved in the calculation, and σ denotes the sigmoid activation function.

3.4. Type-Constrained Graph

To capture relation correlations using entity type constraints from a global perspective of the dataset, we construct a heterogeneous graph structure, called Type-Constrained Graph (TCG), which is used to specify all allowed subject-object entity types for each relation category. Then, the graph is encoded with the graph attention networks (GATs) [19] to generate all relation embeddings, which embody correlation information and are leveraged to further construct additional features for each entity pair that imply correlation knowledge. Next, we present the details of graph construction, graph encoder and additional features construction in turn.

3.4.1. Graph Definition

The Type-Constrained Graph is constructed by performing statistics on the train set, which shows the constraints between relation categories and entity types from the global view of the dataset. Specifically, the types of nodes and edges are defined as follows:

- **Nodes**: The graph involves two types of nodes: relation categories and entity types.
- **Edges**: For connections between nodes, two types of edges are taken into account, i.e., "*subject_type_is*" and "*object_type_is*", which formulate all allowed subject and object entity types of a specific relation category, respectively.

In this way, different relation categories are connected to each other through their common subject types or object types, which implies the type-constrained correlations between relations.

3.4.2. Graph Encoder

To utilize the above graph to learn all correlation-aware relation embeddings, we use the graph attention networks (GATs) [19] to encode the representations of all nodes. Different from the graph convolutional networks [62] that treat all neighbor nodes equally, GATs assign different and appropriate importance scores to neighbor nodes. Specifically, GATs generally consist of multiple stacked attention layers, where each layer transforms the input node representations through attention mechanism and outputs the resulting representations. Since the above graph contains two types of edges, it is necessary to transfer messages for each edge type individually, and to sum the results over all edge types.

Suppose the representations of all nodes are denoted as $V = [v_1, v_2, ..., v_{N_v}]$, $v_i \in \mathbb{R}^{d_h}$ for i from 1 to N_v, which are randomly initialized. Here N_v is the number of nodes. The attention layer can be described as follows,

$$\alpha_{ij} = \mathtt{softmax}(\mathtt{LeakyReLU}(W_{att}^T \cdot [W_i v_i; W_j v_j] + b_{att})) \tag{7}$$

$$V' = \sigma(\sum_{j \in \mathtt{Ne}(i)} \alpha_{ij} W_j v_j) \tag{8}$$

where $\mathtt{softmax}(\cdot)$ and $\mathtt{LeakyReLU}(\cdot)$ are the activation functions, $\sigma(\cdot)$ is the sigmoid function, $\mathtt{Ne}(i)$ indicates the set of all neighbor nodes of node i, and V' denotes the output node representations.

Due to the above learning process may be unstable, multi-head attention technique is usually employed in Eq. 7. Finally, in order to integrate all types of edges, the resulting node representations can be obtained by summing the output on all edge types.

$$V_{res} = [T, R] = \sum_{k \in \mathcal{K}} V'_k \tag{9}$$

where T is the embedding matrix of all entity types, R is the correlation-aware embedding matrix of all relation categories, and $\mathcal{K}$ is the set of all edge types.

3.5. Additional Features Construction

In order to leverage correlation knowledge to guide the classification, for each entity pair (e_s, e_o), we aggregate relation embeddings R to generate the relation-related feature $r_{(s,o)}$,

$$\alpha_{(s,o)} = \texttt{softmax}([h_{e_s}; h_{e_o}] \cdot W_{(s,o)} \cdot R^{\mathsf{T}}) \tag{10}$$

$$r_{(s,o)} = \alpha_{(s,o)} \cdot R \tag{11}$$

where $W_{(s,o)} \in \mathbb{R}^{2d_h \times d_h}$ is the trainable weight matrix. The resulting feature $r_{(s,o)}$ can be considered to contain all correlation information that is required to categorize the entity pair (e_s, e_o). Therefore, $r_{(s,o)}$ can be input into the classifier by modifying Eq. 4 and Eq. 5 as follows:

$$f_s = \tanh\left(W_s h_{e_s} + W_{c_3}[c_{(s,o)}; t_s; r_{(s,o)}]\right) \tag{12}$$

$$f_o = \tanh\left(W_o h_{e_o} + W_{c_4}[c_{(s,o)}; t_o; r_{(s,o)}]\right) \tag{13}$$

where $\{W_{c_3}, W_{c_4}\} \in \mathbb{R}^{2d_h \times d_h}$ are all trainable weights, t_s and t_o are entity type embeddings of subject and object entities. This is one of the most intuitive methods for enhancing classification using relation embeddings R. While other approaches may also be feasible, this intuitive way achieves consistently significant performance improvements in subsequent experiments.

3.6. Type-Constrained Loss

In contrast to the dataset-level global view of TCG, here we consider the local view at the entity pair level, and propose an objective function called Type-Constrained Loss (TCL). Specifically, we argue that "*For each entity pair, the classification probabilities of relations matching its entity* types should be greater than those mismatching*". In other words, relation categories matching its entity types should receive more attention from the classifier, while relation categories mismatching its types are never likely to be expressed. To this end, we define the entity type matching function $\mathcal{M}(e_s, e_o, r)$, which takes 1 when the subject-object entity types match the relation r, and 0 when it does not.

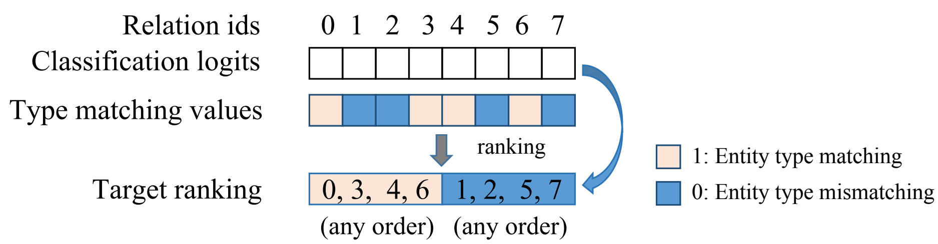


Figure 3: The illustration of Type-Constrained Loss function.

The Type-Constrained Loss is a ranking-based loss [20]. Given an entity pair, for each relation category matching entity types along with probability p_r, it penalizes the cases in which the probability of relation categories mismatching types is greater than p_r. We expect the probabilities of relation categories mismatching entity types to be smaller. For example, in Figure 3, the final target ranking should be that relation categories matching entity types are ranked ahead of those mismatching. The detail definition is as follows,

$$\mathcal{L}_{tcl} = \frac{1}{|D| \times |d| \times |C_{\texttt{match}}|} \sum_{d \in D} \sum_{(e_s, e_o) \in d} \sum_{r \in C_{\texttt{match}}} \frac{\texttt{Mismatch-Rank}(r)}{\texttt{Rank}(r)} \tag{14}$$

$$C_{\texttt{match}} = \{r | r \in \mathcal{R} \;\texttt{\&}\; \mathcal{M}(e_s, e_o, r) = 1\} \tag{15}$$

where D denotes the train set containing all documents, d indicates the group of entity pairs, $C_{\texttt{match}}$ denotes the set of relations matching entity types, $\texttt{Mismatch-Rank}(r)$ is the number of type-mismatched relations with probability greater than p_r, and $\texttt{Rank}(r)$ is the number of all relations with probability greater than p_r. Since each entity pair expresses at most 4 semantic relationships on DocRED and DWIE datasets, this loss can be simplified by retaining the Top-4 type-matched relations for each entity pair, i.e., replacing "$r \in C_{\texttt{match}}$" in Eq. 14 with "$r \in \texttt{TopK}(C_{\texttt{match}})$".

3.7. Training objective

The primary objective for DocRE task is to minimize the binary cross-entropy loss $\mathcal{L}_{re}$, defined as,

$$\mathcal{L}_{re} = \frac{1}{|D| \times |d|} \sum_{d \in D} \sum_{(e_s, e_o) \in d} \sum_{r \in \mathcal{R}} [\mathcal{I}(r) \cdot P(r|e_s, e_o) + (1 - \mathcal{I}(r)) \cdot (1 - P(r|e_s, e_o))] \tag{16}$$

where $\mathcal{I}(r)$ is the ground-truth label of the relation r corresponding to entity pair (e_s, e_o).

Besides, to aggregate the correct relation embeddings for each entity pair in Eq. 11, an additional loss function is proposed to give greater weights to the relation categories expressed by entity pair (e_s, e_o) in Eq. 10, defined as,

$$\mathcal{L}_{aux} = \frac{1}{|D| \times |d|} \sum_{d \in D} \sum_{(e_s, e_o) \in d} \sum_{r \in \mathcal{R}} [\mathcal{I}(r) \cdot \alpha_{(s,o)} + (1 - \mathcal{I}(r)) \cdot (1 - \alpha_{(s,o)})] \tag{17}$$

Finally, the whole loss can be calculated by the harmonic mean operation, as [17],

$$\mathcal{L} = \frac{1 + \beta + \xi}{\frac{1}{\mathcal{L}_{re}} + \frac{\beta}{\mathcal{L}_{aux}} + \frac{\xi}{\mathcal{L}_{tcl}}} \tag{18}$$

where β and ξ are trade-off coefficients.

4. Experiments and Results

In this section, we conduct experiments to compare the proposed DocRE-TCGL approach with other baselines, and further analyze the performance of DocRE-TCGL on long-tail categories and multi-label instances. For the completeness of narrative, the experimental settings are first presented in detail.

4.1. Experimental Settings

4.1.1. Benchmarks

We choose two popular DocRE benchmarks to conduct our experiments, including DocRED [7] and DWIE [21], and display the performance for each setup. For the sake of experimental fairness, two datasets are pre-processed with identical methods as Yao et al. [7] and Ru et al. [63], respectively. The details of benchmarks are displayed in Table 1, including the average number of entities (i.e., **Entities**), the average number of triplet facts (i.e., **Triplet Facts**) and the number of multi-label instances (i.e., **Multi-label Instances**) in each document of train set.

It is observed that, on average, each document in DocRED comprises 19.49 named entities articulating 12.51 triplets. Similarly, on DWIE, the numbers are 27.40 and 23.94. Besides, two datasets involve 2466 and 2880 multi-label entity pairs, respectively, which substantiate the multi-label essence inherent in DocRE task. As for other well-known benchmarks, CDR [64] and GDA [65] involve just a non-`NA` relation category, and are not suitable for capturing correlations between relations.

Table 1: Details of benchmarks exploited in the following experiments.

Benchmarks	Train	Dev	Test	Relations	Entities	Triplet Facts	Multi-label Instances
DocRED	3053	1000	1000	96	19.49	12.51	2466
DWIE	602	98	99	65	27.40	23.94	2880

4.1.2. Metrics

We employ **F1** and **Ign. F1** as metrics for overall performance, here **Ign. F1** indicates F1 score when eliminating the instances existing in the train set. Under long-tail setup, we first compute the F1 value for each relation category, then average the F1 values for all categories with less than K training instances, i.e., the macro-averaged F1 for long-tailed categories (denoted by **Macro@K**), which regards all categories fairly and will not be affected the extreme values. As for multi-label instances, **F1** scores on entity pairs expressing two, three, and four labels, are reported.

4.1.3. Baselines

We select some typical or up-to-date models as baselines for comparison experiments, i.e., CNN [22], LSTM/BiLSTM [26], Context-Aware [66], CorefBERT [67], GAIN [32], SSAN [68], ATLOP [41], ERA/ERACL [69], RSMAN [42], MPCA [70], CPT-RI [71], ChatGPT [13] powered by GPT-3.5 or GPT-4 [14], and DocRE-CoOccur [17]. These baselines are sorted according to the ascending order of their published year. For ChatGPT, we use `gpt-3.5-turbo` and `gpt-4-0125-preview`.

4.1.4. Implementation Details

The proposed DocRE-TCGL is implemented based on the widely-used PyTorch and Transformers [72] libraries. For document encoder, we choose the pre-trained BERT-base-cased [73] and RoBERTa-Large [74]. For optimization, the AdamW optimizer with warmup technique is used during training. Besides, following Han et al. [17], we adopt the identical hyper-parameters including batch size, learning rate, warmup rate, hidden size,etc. The training objective coefficients, β and ξ, are determined on the development set through a search within the range $[0.1, 0.2, ..., 0.9, 1.0, 2.0, 3.0, 4.0, 5.0]$, selecting the values that yield the best F1 score. When performing evaluation, we apply a global threshold to ascertain the existence of category r between entity pair (e_s, e_o). It is selected from the range $[0.1, 0.15, ..., 0.95]$ based on the best F1 score achieved on development set. Our DocRE-TCGL is trained with 1 NVIDIA GeForce RTX 3090 GPU, and it takes about 2.0∼2.5 hours to train 50 epochs.

Table 2: Main hyper-parameters in the training phrase for different datasets.

Benchmarks	DocRED	DWIE
batch_size	4	4
epochs	50	30
learning rate (PLMs/others)	5e-5/1e-4	5e-5/1e-4
warmup rate	6%	6%
β,ξ	4.0,1.0	4.0,2.0

Table 3: The overall performance on DocRED and DWIE datasets. Our DocRE-TCGL model is trained five times by changing random seeds. The baselines' performance on DocRED are from their original publication, while the results on DWIE are from the paper of Yu et al. [42]. All baselines use BERT-Base as the document encoder. Results marked with † symbol indicate that the improvements pass the two-side T-Test($p < 0.05$).

Benchmarks	DocRED				DWIE			
	Dev		Test		Dev		Test	
Models	Ign. F1	F1	Ign. F1	F1	Ign. F1	F1	Ign. F1	F1
CNN [7]	37.99	43.45	36.44	42.33	37.65	47.73	34.65	46.14
LSTM [7]	44.41	50.66	43.60	50.12	40.86	51.77	40.81	52.60
BiLSTM [7]	45.12	50.95	44.73	51.06	40.46	51.92	42.03	54.47
Context-Aware [7]	44.84	51.10	43.93	50.64	42.06	53.05	45.37	56.58
CorefBERT [67]	55.32	57.51	54.54	56.96	57.18	61.42	61.71	66.59
GAIN [32]	59.14	61.22	59.00	61.24	58.63	62.55	62.37	67.57
SSAN [68]	57.04	59.19	56.06	58.41	58.62	64.49	62.58	69.39
ATLOP [41]	59.22	61.09	59.31	61.30	59.03	64.82	62.09	69.94
ERA [69]	59.30	61.30	58.71	60.97	-	-	-	-
ERACL [69]	59.72	61.80	59.08	61.36	-	-	-	-
RSMAN$_{\texttt{SSAN}}$ [42]	57.22	59.25	57.02	59.29	60.02	65.88	63.42	70.95
MPCA [70]	57.93	60.14	57.78	60.24	-	-	-	-
CPT-RI [71]	60.02	62.13	59.92	61.87	-	-	-	-
DocRE-CoOccur [17]	59.39	61.34	59.12	61.32	61.10	65.73	65.64	71.56
ChatGPT$_{\texttt{GPT-3.5}}$ [13]	7.66	7.83	-	-	2.03	2.28	1.35	1.75
ChatGPT$_{\texttt{GPT-4}}$ [14]	15.05	15.42	-	-	13.27	15.14	14.05	16.24
DocRE-Base$_{\texttt{BERT-BASE}}$	58.09$\pm$0.11	60.10$\pm$0.12	58.03	60.20	58.40$\pm$0.26	63.38$\pm$0.33	62.92$\pm$0.64	69.12$\pm$0.56
DocRE-TCGL$_{\texttt{BERT-BASE}}$	**59.41†$\pm$0.10**	**61.27†$\pm$0.07**	**59.23**	**61.17**	**62.03†$\pm$0.45**	**66.80†$\pm$0.36**	**67.01†$\pm$0.56**	**72.82†$\pm$0.54**
	$\uparrow$1.32	$\uparrow$1.17	$\uparrow$1.20	$\uparrow$0.97	$\uparrow$3.63	$\uparrow$3.42	$\uparrow$4.09	$\uparrow$3.70
DocRE-Base$_{\texttt{RoBERTa-LARGE}}$	59.92$\pm$0.39	61.51$\pm$0.38	59.44	61.24	71.82$\pm$0.13	75.35$\pm$0.11	74.94$\pm$0.36	78.94$\pm$0.48
DocRE-TCGL$_{\texttt{RoBERTa-LARGE}}$	**61.32†$\pm$0.16**	**63.08†$\pm$0.28**	**60.94**	**62.79**	**73.07†$\pm$0.63**	**76.63†$\pm$0.49**	**76.52†$\pm$0.53**	**80.55†$\pm$0.44**
	$\uparrow$1.40	$\uparrow$1.57	$\uparrow$1.50	$\uparrow$1.55	$\uparrow$1.25	$\uparrow$1.28	$\uparrow$1.58	$\uparrow$1.61
DocRE-TCGL$_{\texttt{BERT-BASE}}$	**59.41$\pm$0.10**	**61.27$\pm$0.07**	**59.23**	**61.17**	**62.03$\pm$0.45**	**66.80$\pm$0.36**	**67.01$\pm$0.56**	**72.82$\pm$0.54**
w/o TCG	58.73$\pm$0.16	60.72$\pm$0.17	58.74	60.81	59.97$\pm$0.43	65.23$\pm$0.25	65.29$\pm$0.55	71.33$\pm$0.61
w/o TCL	58.96$\pm$0.15	60.92$\pm$0.13	59.18	61.08	61.45$\pm$0.47	65.99$\pm$0.44	66.08$\pm$0.37	71.73$\pm$0.50
w/o TCG and TCL	58.36$\pm$0.10	60.30$\pm$0.09	58.33	60.40	59.05$\pm$0.36	64.44$\pm$0.27	64.43$\pm$0.41	69.74$\pm$0.58
w/o Ent. Type Emb.	58.09$\pm$0.11	60.10$\pm$0.12	58.03	60.20	58.40$\pm$0.26	63.38$\pm$0.33	62.92$\pm$0.64	69.12$\pm$0.56

4.2. Main Results

Table 3 showcases the result comparison of DocRE-TCGL with baselines on two commonly-used benchmarks. We train DocRE-TCGL model five times by changing random seeds, and display all mean values and corresponding standard deviations. Since DocRED's test set does not provide relational labels, and its results must be obtained through CodaLab cite, we do not report the corresponding mean values.

The results reveal that the base model DocRE-Base surpasses several previous BERT-based baseline models, e.g., CorefBERT and SSAN, on both datasets, which indicates that our DocRE-Base yields competitive results. Further, our proposed DocRE-TCGL consistently outperforms DocRE-Base model, due to the utilization of relation correlations. Specifically, in terms of F1 score, it boosts the results of DocRE-Base on F1 score by 1.32, 1.17, 1.20, and 0.97 on the DocRED dataset, and 3.63, 3.42, 4.09, and 3.70 on DWIE. The significant improvements substantiate the robustness and effectiveness of proposed methodology. The improvements are validated by two-sided T-Test with $p < 0.05$. We also report all results of changing document encoder from BERT-base-cased [73] to RoBERTa-Large [74]. It can be found that similar performance is obtained, and the performance improvements are consistent and noticeable.

4.3. Ablation Study

To validate the effectiveness of each module, on both datasets, ablative experiments are conducted, by removing one component at a time. As can be seen from the bottom of Table 3, without TCG or TCL, there is a varied degradation in the model performance (i.e., **w/o TCG** and **w/o TCL**), while without both TCG and TCL, a huge performance drop is observed, and its performance is very close to our base model DocRE-Base (i.e., **w/o TCG and TCL**). When further removing entity type embeddings t_s and t_o in Eqs. 12- 13 in the absence of TCG and TCL (i.e., **w/o Ent. Type Emb.**), the model becomes the same as the base model. As can be observed from the last two lines in Table 3, the performance improvement is marginal when entity type embeddings are utilized alone.

4.4. Further Discussion

4.4.1. Performance on Long-tailed Categories

To investigate how relation correlations influence the result of long-tail categories, we carry out experiments on relation categories with fewer than K training triplets. The macro-averaged F1 scores over these long-tailed categories, denoted as **Macro@K**, are calculated and presented in Table 4.

Due to DocRED dataset does not provide relational labels for the test set, following [17], we just report the results of development set on two datasets. For DocRED dataset, we set K to 500, 200, and 100. For DWIE dataset, we set K to 100 and 50. The column "**Macro@all**" refers to the macro-averaged F1 score across all categories, regardless of whether it is long-tailed. The results of baselines are either provided by their original paper or derived from their official codes. All models implemented by us are trained five times by changing random seeds, and the mean values and standard deviation values are reported.

Our proposed DocRE-TCGL consistently outperforms most baseline models on long-tail categories for both benchmarks. The less the training instances, the more significant the performance gains. It is worth noting that

Table 4: Performance on long-tailed relations. DocRE-TCGL is trained five times by changing random seeds, and the mean values and standard deviations on the development set are reported. Since DocRED's test set is not publicly accessible, its performance could not be exhibited. The results marked by ‡ are from the paper of Du et al. [69]. Other results are produced by our implementation based on their codes. † indicates that the improvements pass the two-sided T-Test ($p < 0.05$).

Benchmarks	DocRED				DWIE		
Models	Macro@all	Macro@500	Macro@200	Macro@100	Macro@all	Macro@100	Macro@50
CorefBERT [67]	36.32±0.31	32.07±0.29	24.69±0.35	17.12±0.35	27.60±0.84	9.19±1.10	4.78±0.83
GAIN [32]	38.47±0.24	33.99±0.28	26.29±0.33	18.40±0.54	30.88±0.74	10.55±0.95	6.84±0.85
SSAN [68]	36.82±0.63	32.39±0.71	24.78±0.70	18.23±0.80	21.42±0.84	6.49±1.39	2.25±1.41
ATLOP [41]	39.24±0.30	34.85±0.36	26.63±0.41	18.68±0.47	30.96±0.56	11.91±0.16	7.10±0.39
ERA‡ [69]	40.55	36.21	28.51	20.50	-	-	-
ERACL‡ [69]	41.34	37.13	29.43	22.31	-	-	-
RSMAN$_{\texttt{SSAN}}$ [42]	35.82±0.47	31.40±0.57	23.63±0.50	17.19±0.88	22.35±0.62	6.79±0.35	2.62±0.31
BERT-CoOccur [17]	40.81±0.35	36.55±0.40	28.76±0.63	21.38±0.96	32.80±1.25	13.02±1.46	8.59±1.73
ChatGPT$_{\texttt{GPT-3.5}}$ [13]	6.95	6.47	4.22	2.89	1.59	2.29	2.47
ChatGPT$_{\texttt{GPT-4}}$ [14]	13.97	11.76	8.00	7.24	2.44	3.52	3.27
DocRE-Base$_{\texttt{BERT-BASE}}$	39.70±0.47	35.35±0.59	27.66±0.71	19.84±0.64	28.17±0.40	6.53±0.55	2.47±0.50
DocRE-TCGL$_{\texttt{BERT-BASE}}$	**40.71†±0.35**	**36.48†±0.41**	**29.05†±0.61**	**22.05†±0.57**	**32.93†±1.03**	**12.60†±1.00**	**8.73†±0.58**
	↑1.01	↑1.13	↑1.39	↑2.21	↑4.76	↑6.07	↑6.26
DocRE-Base$_{\texttt{RoBERTa-LARGE}}$	41.42±0.46	37.25±0.26	29.19±0.30	21.99±0.22	40.19±0.65	17.09±0.66	13.17±0.55
DocRE-TCGL$_{\texttt{RoBERTa-LARGE}}$	**42.30±0.48**	**38.64±0.36**	**30.50±0.35**	**23.18±0.04**	**42.99±1.13**	**21.61±0.61**	**17.47±1.26**
	↑0.88	↑1.39	↑1.31	↑1.19	↑2.80	↑4.52	↑4.30
DocRE-TCGL$_{\texttt{BERT-BASE}}$	**40.71±0.35**	**36.48±0.41**	**29.05±0.61**	**22.05±0.57**	**32.93±1.03**	**12.60±1.00**	**8.73±0.58**
w/o TCG	40.18±0.27	35.97±0.16	27.99±0.35	21.12±0.77	31.16±0.30	10.97±0.47	6.82±0.67
w/o TCL	40.30±0.42	36.07±0.49	28.19±0.50	20.61±0.52	31.64±0.26	11.80±0.77	7.08±0.34
w/o TCG and TCL	39.87±0.36	35.46±0.35	27.79±0.49	20.05±0.33	29.98±0.39	8.29±0.33	5.46±0.47
w/o Ent. Type Emb.	39.70±0.47	35.35±0.59	27.66±0.71	19.84±0.64	28.17±0.40	6.53±0.55	2.47±0.50

Macro@K values are improved by up to 6.26 and 2.21, on two datasets, respectively. Additionally, for the baselines tailor-made for long-tail problem, including ERA, ERACL and BERT-CoOccur, our proposed DocRE-TCGL achieves competitive performance with them. These confirm that relation correlations have great potential in tackling long-tail relations.

4.4.2. Performance on Multi-Label Instances

To investigate the influence of relation correlations on multi-label instances, we evaluate DocRE-TCGL model on all multi-label instances from the development set. Some of these entity pairs even express four semantic relationships simultaneously. Due to the unavailability of labels for the DocRED test set, following Han et al. [17], we just report F1 scores on the development set for both datasets in Table 5. Note that, each label of an instance is independently evaluated, if an instance has two labels, then two triplet facts it contained need to be judged respectively. We train all models five times by changing random seeds, display all mean values and standard deviations. Since ERA and ERACL lack open-source codes, their performance cannot be reported.

We can observe that relation correlations lead to consistent improvements in handling multi-label entity pairs across both datasets, effectively alleviating the multi-label problem. The more relational labels, the more significant the performance improvement. Compared to our base model DocRE-Base, in terms of F1 score, DocRE-TCGL model improves up to 5.74 and 4.91 on the DocRED and DWIE datasets, respectively. In addition, compared to the DocRE-CoOccur model, which is specialized in solving the multi-label problem, our DocRE-TCGL model achieves competitive performance with it, confirming that type-constrained correlations can substantially mitigate the multi-label problem in DocRE task.

Table 5: Performance on multi-label instances. DocRE-TCGL is trained five times by changing random seeds, and all mean values and standard deviations on the development set are reported. The results marked by † indicates that the improvements pass the two-sided T-Test ($p < 0.05$).

Benchmarks	DocRED				DWIE		
Models	Two	Three	Four	Mean	Two	Three	Mean
CorefBERT [67]	67.55±0.69	50.38±1.29	31.48±1.85	49.80±1.04	66.17±0.95	78.29±0.49	72.23±0.54
GAIN [32]	67.43±0.32	47.72±0.67	40.00±0.00	51.72±0.33	67.06±0.65	77.20±0.61	72.13±0.43
SSAN [68]	67.06±0.71	48.75±1.74	40.00±0.00	51.94±0.57	54.43±1.40	70.47±1.51	62.45±1.44
ATLOP [41]	69.06±0.72	50.60±1.46	38.16±2.98	52.60±1.42	72.72±0.35	77.83±1.38	75.27±0.73
RSMAN [42]	68.13±0.68	53.37±0.64	45.53±1.26	55.67±0.25	58.26±0.85	72.15±0.31	65.20±0.29
DocRE-CoOccur [17]	70.97±0.73	55.16±0.58	46.65±1.95	57.59±0.63	73.98±1.63	79.29±1.41	76.63±0.86
ChatGPT$_{\texttt{GPT-3.5}}$ [13]	8.93	5.33	22.22	12.16	2.53	0.70	1.62
ChatGPT$_{\texttt{GPT-4}}$ [14]	17.73	15.19	45.46	26.13	19.75	0.62	10.19
DocRE-Base$_{\texttt{BERT-BASE}}$	69.80±0.69	49.56±1.49	34.26±1.62	51.21±1.17	70.13±0.59	77.11±1.03	73.62±0.38
DocRE-TCGL$_{\texttt{BERT-BASE}}$	**70.53†±0.35**	**53.57†±0.92**	**40.00†±0.00**	**54.70†±0.38**	**74.12†±1.06**	**82.02†±1.12**	**78.07†±0.85**
	↑0.73	↑4.01	↑5.74	↑3.49	↑3.99	↑4.91	↑4.45
DocRE-Base$_{\texttt{RoBERTa-LARGE}}$	71.06±0.29	53.50±0.78	45.09±1.11	56.55±0.12	80.12±0.86	86.38±0.62	83.25±0.67
DocRE-TCGL$_{\texttt{RoBERTa-LARGE}}$	**71.93±0.19**	**55.75±1.82**	**47.62±0.00**	**58.43±0.57**	**81.99±0.90**	**88.11±0.69**	**85.05±0.70**
	↑0.87	↑2.25	↑2.53	↑1.88	↑1.87	↑1.73	↑1.80
DocRE-TCGL$_{\texttt{BERT-BASE}}$	**70.53±0.35**	**53.57±0.92**	**40.00±0.00**	**54.70±0.38**	**74.12±1.06**	**82.02±1.12**	**78.07±0.85**
w/o TCG	70.19±0.73	53.58±1.56	39.88±0.84	54.55±0.59	72.16±0.56	80.75±0.66	76.45±0.49
w/o TCL	70.39±0.72	54.32±0.25	41.52±1.05	55.41±0.97	72.91±0.88	81.04±0.61	76.97±0.71
w/o TCG and TCL	69.90±0.44	51.36±0.60	40.00±0.00	53.75±0.30	70.62±0.64	78.47±0.55	74.83±0.53
w/o Ent. Type Emb.	69.80±0.69	49.56±1.49	34.26±1.62	51.21±1.17	70.13±0.59	77.11±1.03	73.62±0.38

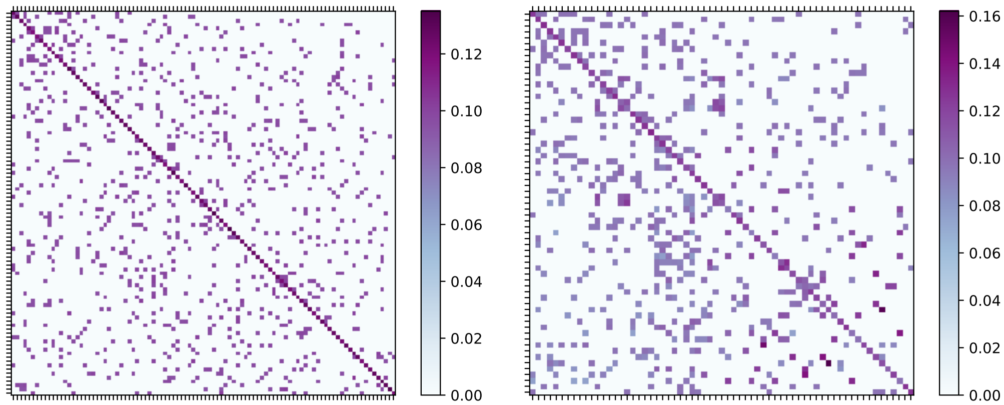


Figure 4: Visualization of correlations. Left figure is for DocRED, right figure is for DWIE.

4.4.3. Visualization of Relation Correlations

To verify that the proposed model indeed captures a large number of type-constrained relation correlations, we perform a visualization here. Specifically, we use the relation embeddings learned by Section 3.4.2 on both DocRED and DWIE datasets for visualization. For measuring the degree of relatedness between relation categories, the dot-product operation is employed to compute the similarity matrix for all categories, which is exhibits in Figure 4. Other methods of calculating the relation similarity matrix are also feasible, here the dot-product is used because of its simplicity.

In Figure 4, for easy observation, we just keep and show the top-10 correlated categories in each row. We can find that our proposed DocRE-TCGL indeed learns a great deal of type-constrained relation correlations on both datasets, confirming the effectiveness of Type-Constrained Graph and Type-Constrained Loss. These correlation knowledges can enhance existing DocRE models to alleviate the long-tail and multi-label problems.

5. Conclusion

Document-level relation extraction task has two performance bottlenecks, i.e., the long-tail challenge and multi-label challenge. In this paper, we aim to enhance DocRE models using relation correlations to alleviate the above two issues at the same time, and propose a methodology for modeling relation correlations with entity type constraints, from both global and local perspectives. The entity type constraints are binding relationships between relation categories and subject-object entity types, i.e., all allowed entity types for a specific relation category are fixed.

Specifically, from the global perspective, we perform statistics on the train set, and construct the Type-Constrained Graph (TCG) to formulate all possible subject/object types for each relation category. The correlations exist between relations with the same subject-object type. Then, the Graph Attention Networks (GATs) with multi-head attention mechanism is used to encode all relation embeddings. These embeddings contain massive correlation knowledge, and then are exploited to generate additional features for each entity pair in order to guide the classification. From the local perspective, we argue that *given an entity pair, the classification probabilities of relations matching its entity types should be greater than those unmatched*, and propose a ranking-based Type-Constrained Loss (TCL) to make the matched relations have greater probabilities. Extensive experiments on two commonly-used benchmarks, including DocRED and DWIE, are carried out. The results reveal that the proposed DocRE-TCGL obtains consistent performance improvements, and significantly outperforms the typical or up-to-date baselines on both long-tailed and multi-label setups, confirming the great potential of relation correlations.

CRediT authorship contribution statement

Ridong Han: Conceptualization, Data Curation, Methodology, Software, Writing-Original Draft, Visualization. **Tao Peng**: Project Administration, Resources, Funding Acquisition. **Beibei Zhu**: Software, Validation. **Haijia Bi**: Data Curation. **Jiayu Han**: Conceptualization, Validation. **Xinzheng Xu**: Validation. **Lu Liu**: Funding Acquisition, Writing-Review & Editing.

Declaration of Competing Interest

The authors confirm that there are no known conflicts of interest or personal relationships that might have biased the work presented in this article.

Acknowledgements

Our sincere thanks go to the anonymous reviewers for their valuable time and effort in scrutinizing this manuscript.

This paper is funded by the National Natural Science Foundation of China under grant No.61872163 and 61806084, and Jilin Province Key Scientific and Technological Research and Development Project under grant No.20210201131GX.

References

[1] C. N. dos Santos, B. Xiang, B. 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, pp. 626–634. `doi:10.3115/v1/p15-1061`.
URL `https://doi.org/10.3115/v1/p15-1061`

[2] Y. Lin, S. Shen, Z. Liu, H. Luan, M. Sun, Neural relation extraction with selective attention over instances, in: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL, 2016, pp. 2124–2133. `doi:10.18653/v1/p16-1200`.
URL `https://doi.org/10.18653/v1/p16-1200`

[3] J. Qu, D. Ouyang, W. Hua, Y. Ye, X. Li, Distant supervision for neural relation extraction integrated with word attention and property features, Neural Networks 100 (2018) 59–69. `doi:10.1016/j.neunet.2018.01.006`.
URL `https://doi.org/10.1016/j.neunet.2018.01.006`

[4] T. Peng, R. Han, H. Cui, L. Yue, J. Han, L. Liu, Distantly supervised relation extraction using global hierarchy embeddings and local probability constraints, Knowledge-Based Systems 235 (2022) 107637. `doi:10.1016/j.knosys.2021.107637`.
URL `https://doi.org/10.1016/j.knosys.2021.107637`

[5] K. Zhou, Q. Qiao, Y. Li, Q. Li, Improving distantly supervised relation extraction by natural language inference, in: Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI, 2023, pp. 14047–14055. doi:10.1609/AAAI.V37I11.26644.
URL https://doi.org/10.1609/aaai.v37i11.26644

[6] Y. Shang, H. Huang, X. Sun, W. Wei, X. Mao, Learning relation ties with a force-directed graph in distant supervised relation extraction, ACM Transactions on Information Systems 41 (1) (2023) 10:1–10:23. doi:10.1145/3520082.
URL https://doi.org/10.1145/3520082

[7] Y. Yao, D. Ye, P. Li, X. Han, Y. Lin, Z. Liu, Z. Liu, L. Huang, J. Zhou, M. Sun, Docred: A large-scale document-level relation extraction dataset, in: Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL, 2019, pp. 764–777. doi:10.18653/v1/p19-1074.
URL https://doi.org/10.18653/v1/p19-1074

[8] C. Yuan, H. Huang, C. Feng, G. Shi, X. Wei, Document-level relation extraction with entity-selection attention, Information Sciences 568 (2021) 163–174. doi:10.1016/J.INS.2021.04.007.
URL https://doi.org/10.1016/j.ins.2021.04.007

[9] D. Zeng, J. Zhu, H. Chen, J. Dai, L. Jiang, Document-level denoising relation extraction with false-negative mining and reinforced positive-class knowledge distillation, Information Processing & Management 61 (1)

(2024) 103533. doi:10.1016/J.IPM.2023.103533.
URL https://doi.org/10.1016/j.ipm.2023.103533

[10] L. Li, R. Lian, H. Lu, Document-level biomedical relation extraction with generative adversarial network and dual-attention multi-instance learning, in: Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine, BIBM, 2021, pp. 438–443. doi:10.1109/BIBM52615.2021.9669590.
URL https://doi.org/10.1109/BIBM52615.2021.9669590

[11] W. Xu, K. Chen, L. Mou, T. Zhao, Document-level relation extraction with sentences importance estimation and focusing, in: Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL, 2022, pp. 2920–2929. doi:10.18653/V1/2022.NAACL-MAIN.212.
URL https://doi.org/10.18653/v1/2022.naacl-main.212

[12] R. Zhang, Y. Li, L. Zou, A novel table-to-graph generation approach for document-level joint entity and relation extraction, in: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics, ACL, 2023, pp. 10853–10865. doi:10.18653/V1/2023.ACL-LONG.607.
URL https://doi.org/10.18653/v1/2023.acl-long.607

[13] OpenAI, Introducing chatgpt, OpenAI Blog (2023).
URL https://openai.com/blog/chatgpt

[14] OpenAI, GPT-4 technical report, CoRR abs/2303.08774 (2023). arXiv:2303.08774, doi:10.48550/arXiv.2303.08774.
URL https://doi.org/10.48550/arXiv.2303.08774

[15] B. Li, G. Fang, Y. Yang, Q. Wang, W. Ye, W. Zhao, S. Zhang, Evaluating chatgpt's information extraction capabilities: An assessment of performance, explainability, calibration, and faithfulness, CoRR abs/2304.11633 (2023). arXiv:2304.11633, doi:10.48550/arXiv.2304.11633.
URL https://doi.org/10.48550/arXiv.2304.11633

[16] R. Han, T. Peng, C. Yang, B. Wang, L. Liu, X. Wan, Is information extraction solved by chatgpt? an analysis of performance, evaluation criteria, robustness and errors, CoRR abs/2305.14450 (2023). arXiv:2305.14450, doi:10.48550/ARXIV.2305.14450.
URL https://doi.org/10.48550/arXiv.2305.14450

[17] R. Han, T. Peng, B. Wang, L. Liu, P. Tiwari, X. Wan, Document-level relation extraction with relation correlations, Neural Networks 171 (2024) 14–24. doi:10.1016/j.neunet.2023.11.062.
URL https://doi.org/10.1016/j.neunet.2023.11.062

[18] Y. Huang, Z. Lin, Document-level relation extraction with relation correlation enhancement, in: Proceedings of the 30th International Conference of Neural Information Processing, ICONIP, Vol. 1967, 2023, pp. 427–440. doi:10.1007/978-981-99-8178-6_33.
URL https://doi.org/10.1007/978-981-99-8178-6_33

[19] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, Y. Bengio, Graph attention networks, in: Proceedings of the 6th International Conference on Learning Representations, ICLR, 2018.
URL https://openreview.net/forum?id=rJXMpikCZ

[20] K. Oksuz, B. C. Cam, E. Akbas, S. Kalkan, A ranking-based, balanced loss function unifying classification and localisation in object detection, in: Advances in Neural Information Processing Systems, NeurIPS, 2020.
URL https://proceedings.neurips.cc/paper/2020/hash/b2eeb7362ef83deff5c7813a67e14f0a-Abstract.html

[21] K. Zaporojets, J. Deleu, C. Develder, T. Demeester, DWIE: an entity-centric dataset for multi-task document-level information extraction, Information Processing & Management 58 (4) (2021) 102563. doi:10.1016/j.ipm.2021.102563.
URL https://doi.org/10.1016/j.ipm.2021.102563

[22] D. Zeng, K. Liu, S. Lai, G. Zhou, J. Zhao, Relation classification via convolutional deep neural network, in: Proceedings of the 25th International Conference on Computational Linguistics, COLING, 2014, pp. 2335–2344.
URL https://aclanthology.org/C14-1220/

[23] D. Wang, P. Tiwari, S. Garg, H. Zhu, P. Bruza, Structural block driven enhanced convolutional neural representation for relation extraction, Applied Soft Computing 86 (2020). doi:10.1016/J.ASOC.2019.

105913.
URL https://doi.org/10.1016/j.asoc.2019.105913

[24] R. Han, T. Peng, J. Han, H. Cui, L. Liu, Distantly supervised relation extraction via recursive hierarchy-interactive attention and entity-order perception, Neural Networks 152 (2022) 191–200. doi:10.1016/j.neunet.2022.04.019.
URL https://doi.org/10.1016/j.neunet.2022.04.019

[25] S. Zhang, D. Zheng, X. Hu, M. Yang, Bidirectional long short-term memory networks for relation classification, in: Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation, PACLIC, 2015, pp. 73–78.
URL https://aclanthology.org/Y15-1009/

[26] R. Cai, X. Zhang, H. Wang, Bidirectional recurrent convolutional neural network for relation classification, in: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL, 2016, pp. 756–765. doi:10.18653/v1/p16-1072.
URL https://doi.org/10.18653/v1/p16-1072

[27] Y. Zhang, P. Qi, C. D. Manning, Graph convolution over pruned dependency trees improves relation extraction, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP, 2018, pp. 2205–2215. doi:10.18653/v1/d18-1244.
URL https://doi.org/10.18653/v1/d18-1244

[28] Y. Yuan, L. Liu, S. Tang, Z. Zhang, Y. Zhuang, S. Pu, F. Wu, X. Ren,

Cross-relation cross-bag attention for distantly-supervised relation extraction, in: Proceedings of The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI, 2019, pp. 419–426. doi:10.1609/aaai.v33i01.3301419.
URL https://doi.org/10.1609/aaai.v33i01.3301419

[29] Z. Ye, Z. Ling, Distant supervision relation extraction with intra-bag and inter-bag attentions, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT, 2019, pp. 2810–2819. doi:10.18653/v1/n19-1288.
URL https://doi.org/10.18653/v1/n19-1288

[30] T. Chen, L. Zhou, N. Wang, X. Chen, Joint entity and relation extraction with position-aware attention and relation embedding, Applied Soft Computing 119 (2022) 108604. doi:10.1016/J.ASOC.2022.108604.
URL https://doi.org/10.1016/j.asoc.2022.108604

[31] H. Peng, T. Gao, X. Han, Y. Lin, P. Li, Z. Liu, M. Sun, J. Zhou, Learning from context or names? an empirical study on neural relation extraction, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP, 2020, pp. 3661–3672. doi:10.18653/v1/2020.emnlp-main.298.
URL https://doi.org/10.18653/v1/2020.emnlp-main.298

[32] S. Zeng, R. Xu, B. Chang, L. Li, Double graph based reasoning for document-level relation extraction, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP,

2020, pp. 1630–1640. doi:10.18653/v1/2020.emnlp-main.127.
URL https://doi.org/10.18653/v1/2020.emnlp-main.127

[33] L. Li, R. Lian, H. Lu, J. Tang, Document-level biomedical relation extraction based on multi-dimensional fusion information and multi-granularity logical reasoning, in: Proceedings of the 29th International Conference on Computational Linguistics, COLING, 2022, pp. 2098–2107.
URL https://aclanthology.org/2022.coling-1.183

[34] H. Huang, C. Yuan, Q. Liu, Y. Cao, Document-level relation extraction via separate relation representation and logical reasoning, ACM Transactions on Information Systems 42 (1) (2024) 22:1–22:24. doi:10.1145/3597610.
URL https://doi.org/10.1145/3597610

[35] I. J. Goodfellow, Y. Bengio, A. C. Courville, Deep Learning, Adaptive Computation and Machine Learning, MIT press, 2016.
URL http://www.deeplearningbook.org/

[36] M. Schuster, K. K. Paliwal, Bidirectional recurrent neural networks, IEEE transactions on Signal Processing 45 (11) (1997) 2673–2681. doi:10.1109/78.650093.
URL https://doi.org/10.1109/78.650093

[37] G. Nan, Z. Guo, I. Sekulic, W. Lu, Reasoning with latent structure refinement for document-level relation extraction, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics,

ACL, 2020, pp. 1546–1557. doi:10.18653/v1/2020.acl-main.141.
URL https://doi.org/10.18653/v1/2020.acl-main.141

[38] W. Xu, K. Chen, T. Zhao, Discriminative reasoning for document-level relation extraction, in: Findings of the Association for Computational Linguistics: ACL/IJCNLP, 2021, pp. 1653–1663. doi:10.18653/v1/2021.findings-acl.144.
URL https://doi.org/10.18653/v1/2021.findings-acl.144

[39] W. Xu, K. Chen, T. Zhao, Document-level relation extraction with reconstruction, in: Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI, 2021, pp. 14167–14175. doi:10.1609/AAAI.V35I16.17667.
URL https://ojs.aaai.org/index.php/AAAI/article/view/17667

[40] X. Peng, C. Zhang, K. Xu, Document-level relation extraction via subgraph reasoning, in: Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI, 2022, pp. 4331–4337. doi:10.24963/ijcai.2022/601.
URL https://doi.org/10.24963/ijcai.2022/601

[41] W. Zhou, K. Huang, T. Ma, J. Huang, Document-level relation extraction with adaptive thresholding and localized context pooling, in: Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI, 2021, pp. 14612–14620. doi:10.1609/AAAI.V35I16.17717.
URL https://ojs.aaai.org/index.php/AAAI/article/view/17717

[42] J. Yu, D. Yang, S. Tian, Relation-specific attentions over entity mentions

for enhanced document-level relation extraction, in: Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL, 2022, pp. 1523–1529. doi:10.18653/v1/2022.naacl-main.109.
URL https://doi.org/10.18653/v1/2022.naacl-main.109

[43] Y. Xie, J. Shen, S. Li, Y. Mao, J. Han, Eider: Empowering document-level relation extraction with efficient evidence extraction and inference-stage fusion, in: Findings of the Association for Computational Linguistics: ACL, 2022, pp. 257–268. doi:10.18653/v1/2022.findings-acl.23.
URL https://doi.org/10.18653/v1/2022.findings-acl.23

[44] Q. Tan, R. He, L. Bing, H. T. Ng, Document-level relation extraction with adaptive focal loss and knowledge distillation, in: Findings of the Association for Computational Linguistics: ACL, 2022, pp. 1672–1681. doi:10.18653/v1/2022.findings-acl.132.
URL https://doi.org/10.18653/v1/2022.findings-acl.132

[45] Y. Zhou, W. S. Lee, None class ranking loss for document-level relation extraction, in: Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI, 2022, pp. 4538–4544. doi:10.24963/ijcai.2022/630.
URL https://doi.org/10.24963/ijcai.2022/630

[46] M. Zhang, Z. Zhou, A review on multi-label learning algorithms, IEEE transactions on knowledge and data engineering 26 (8) (2014) 1819–

1837. doi:10.1109/TKDE.2013.39.
URL https://doi.org/10.1109/TKDE.2013.39

[47] Y. Zhu, J. T. Kwok, Z. Zhou, Multi-label learning with global and local label correlation, IEEE transactions on knowledge and data engineering 30 (6) (2018) 1081–1094. doi:10.1109/TKDE.2017.2785795.
URL https://doi.org/10.1109/TKDE.2017.2785795

[48] Z. Jin, Y. Yang, X. Qiu, Z. Zhang, Relation of the relations: A new paradigm of the relation extraction problem, CoRR abs/2006.03719 (2020). arXiv:2006.03719.
URL https://arxiv.org/abs/2006.03719

[49] L. Fu, R. Grishman, Learning relatedness between types with prototypes for relation extraction, in: Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, 2021, pp. 2011–2016. doi:10.18653/v1/2021.eacl-main.172.
URL https://doi.org/10.18653/v1/2021.eacl-main.172

[50] G. Zhou, J. Su, J. Zhang, M. Zhang, Exploring various knowledge in relation extraction, in: Proceedings of the 43rd annual meeting of the association for computational linguistics, ACL, 2005, pp. 427–434.

[51] G. Angeli, M. J. J. Premkumar, C. D. Manning, Leveraging linguistic structure for open domain information extraction, in: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics,

ACL, 2015, pp. 344–354. doi:10.3115/V1/P15-1034.
URL https://doi.org/10.3115/v1/p15-1034

[52] T. T. Tran, P. Le, S. Ananiadou, Revisiting unsupervised relation extraction, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL, 2020, pp. 7498–7505. doi:10.18653/V1/2020.ACL-MAIN.669.
URL https://doi.org/10.18653/v1/2020.acl-main.669

[53] L. Bai, X. Jin, C. Zhuang, X. Cheng, Entity type enhanced neural model for distantly supervised relation extraction (student abstract), in: Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI, 2020, pp. 13751–13752. doi:10.1609/AAAI.V34I10.7147.
URL https://doi.org/10.1609/aaai.v34i10.7147

[54] A. Papaluca, D. Krefl, H. Suominen, A. Lenskiy, Pretrained knowledge base embeddings for improved sentential relation extraction, in: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, ACL, 2022, pp. 373–382. doi:10.18653/V1/2022.ACL-SRW.29.
URL https://doi.org/10.18653/v1/2022.acl-srw.29

[55] Y. Wang, C. Sun, Y. Wu, H. Zhou, L. Li, J. Yan, Unire: A unified label space for entity relation extraction, in: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics, ACL, 2021, pp. 220–231. doi:10.18653/V1/2021.ACL-LONG.19.
URL https://doi.org/10.18653/v1/2021.acl-long.19

[56] Z. Chen, C. Guo, A pattern-first pipeline approach for entity and relation extraction, Neurocomputing 494 (2022) 182–191. doi:10.1016/J.NEUCOM.2022.04.059.
URL https://doi.org/10.1016/j.neucom.2022.04.059

[57] Y. Wu, Y. Chen, Y. Qin, R. Tang, Q. Zheng, A recollect-tuning method for entity and relation extraction, Expert Systems with Applications (2023) 123000.

[58] X. Hu, Z. Hong, C. Zhang, I. King, P. Yu, Think rationally about what you see: Continuous rationale extraction for relation extraction, in: Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR, 2023, pp. 2436–2440.

[59] H. Wang, C. Focke, R. Sylvester, N. Mishra, W. Y. Wang, Fine-tune bert for docred with two-step process, CoRR abs/1909.11898 (2019). arXiv:1909.11898.
URL http://arxiv.org/abs/1909.11898

[60] Y. Zhang, V. Zhong, D. Chen, G. Angeli, C. 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, pp. 35–45. doi:10.18653/v1/d17-1004.
URL https://doi.org/10.18653/v1/d17-1004

[61] L. B. Soares, N. FitzGerald, J. Ling, T. Kwiatkowski, Matching the blanks: Distributional similarity for relation learning, in: Proceedings of the 57th Conference of the Association for Computational Linguistics,

ACL, 2019, pp. 2895–2905. doi:10.18653/v1/p19-1279.
URL https://doi.org/10.18653/v1/p19-1279

[62] T. N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, in: Proceedings of the 5th International Conference on Learning Representations, ICLR, 2017.
URL https://openreview.net/forum?id=SJU4ayYgl

[63] D. Ru, C. Sun, J. Feng, L. Qiu, H. Zhou, W. Zhang, Y. Yu, L. Li, Learning logic rules for document-level relation extraction, in: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP, 2021, pp. 1239–1250. doi:10.18653/v1/2021.emnlp-main.95.
URL https://doi.org/10.18653/v1/2021.emnlp-main.95

[64] J. Li, Y. Sun, R. J. Johnson, D. Sciaky, C. Wei, R. Leaman, A. P. Davis, C. J. Mattingly, T. C. Wiegers, Z. Lu, Biocreative V CDR task corpus: A resource for chemical disease relation extraction, Database J. Biol. Databases Curation 2016 (2016). doi:10.1093/database/baw068.
URL https://doi.org/10.1093/database/baw068

[65] Y. Wu, R. Luo, H. C. M. Leung, H. Ting, T. W. Lam, RENET: A deep learning approach for extracting gene-disease associations from literature, in: Research in Computational Molecular Biology - 23rd Annual International Conference, RECOMB, Vol. 11467 of Lecture Notes in Computer Science, 2019, pp. 272–284. doi:10.1007/978-3-030-17083-7_17.
URL https://doi.org/10.1007/978-3-030-17083-7_17

[66] D. Sorokin, I. Gurevych, Context-aware representations for knowledge base relation extraction, in: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP, 2017, pp. 1784–1789. doi:10.18653/v1/d17-1188.
URL https://doi.org/10.18653/v1/d17-1188

[67] D. Ye, Y. Lin, J. Du, Z. Liu, P. Li, M. Sun, Z. Liu, Coreferential reasoning learning for language representation, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP, 2020, pp. 7170–7186. doi:10.18653/v1/2020.emnlp-main.582.
URL https://doi.org/10.18653/v1/2020.emnlp-main.582

[68] B. Xu, Q. Wang, Y. Lyu, Y. Zhu, Z. Mao, Entity structure within and throughout: Modeling mention dependencies for document-level relation extraction, in: Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI, 2021, pp. 14149–14157. doi:10.1609/AAAI.V35I16.17665.
URL https://ojs.aaai.org/index.php/AAAI/article/view/17665

[69] Y. Du, T. Ma, L. Wu, Y. Wu, X. Zhang, B. Long, S. Ji, Improving long tailed document-level relation extraction via easy relation augmentation and contrastive learning, CoRR abs/2205.10511 (2022). arXiv:2205.10511, doi:10.48550/arXiv.2205.10511.
URL https://doi.org/10.48550/arXiv.2205.10511

[70] X. Ding, G. Zhou, T. Zhu, Multi-perspective context aggregation for document-level relation extraction, Applied Intelligence 53 (6) (2023)

6926–6935. doi:10.1007/S10489-022-03731-W.
URL https://doi.org/10.1007/s10489-022-03731-w

[71] C. Yuan, Y. Cao, H. Huang, Collective prompt tuning with relation inference for document-level relation extraction, Information Processing & Management 60 (5) (2023) 103451. doi:10.1016/J.IPM.2023.103451.
URL https://doi.org/10.1016/j.ipm.2023.103451

[72] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Brew, Huggingface's transformers: State-of-the-art natural language processing, CoRR abs/1910.03771 (2019). arXiv:1910.03771.
URL http://arxiv.org/abs/1910.03771

[73] J. Devlin, M. Chang, K. Lee, K. 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, pp. 4171–4186. doi:10.18653/v1/n19-1423.
URL https://doi.org/10.18653/v1/n19-1423

[74] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized BERT pretraining approach, CoRR abs/1907.11692 (2019). arXiv:1907.11692.
URL http://arxiv.org/abs/1907.11692