Document-level Relation Extraction as Semantic Segmentation

Ningyu Zhang[1,2 *], **Xiang Chen**[1,2 *], **Xin Xie**[1,2], **Shumin Deng**[1,2], **Chuanqi Tan**[3], **Mosha Chen**[3], **Fei Huang**[3], **Luo Si**[3], **Huajun Chen**[1,2 †]
[1] Zhejiang University & AZFT Joint Lab for Knowledge Engine
[2] Hangzhou Innovation Center, Zhejiang University
[3] Alibaba Group
{zhangningyu,xiang_chen,xx2020,231sm,huajunsir}@zju.edu.cn
{chuanqi.tcq,chenmosha.cms,f.huang,luo.si}@alibaba-inc.com

Abstract

Document-level relation extraction aims to extract relations among multiple entity pairs from a document. Previously proposed graph-based or transformer-based models utilize the entities independently, regardless of global information among relational triples. This paper approaches the problem by predicting an entity-level relation matrix to capture local and global information, parallel to the semantic segmentation task in computer vision. Herein, we propose a Document U-shaped Network for document-level relation extraction. Specifically, we leverage an encoder module to capture the context information of entities and a U-shaped segmentation module over the image-style feature map to capture global interdependency among triples. Experimental results show that our approach can obtain state-of-the-art performance on three benchmark datasets DocRED, CDR, and GDA[1].

1 Introduction

Relation extraction (RE) is an important task in the field of information extraction, which has widespread applications [Zhang *et al.*, 2021b; Zhang *et al.*, 2021a]. Previous works [Zeng *et al.*, 2015; Feng *et al.*, 2018] focused on identifying relations within a single sentence, which failed to recognize relations between entities across sentences. However, many relations are expressed over multiple sentences in real-world applications. According to [Yao *et al.*, 2019], above 40.7% of relations can only be identified at the document level. Therefore, it is crucial for models to be able to extract document-level relations.

Recent studies [Yao *et al.*, 2019; Tang *et al.*, 2020; Zeng *et al.*, 2020; Wang *et al.*, 2020a; Zhou *et al.*, 2021] have extended sentence-level RE to the document level. Compared with sentence-level RE that only contains one entity pair to classify in a sentence, document-level RE requires the model

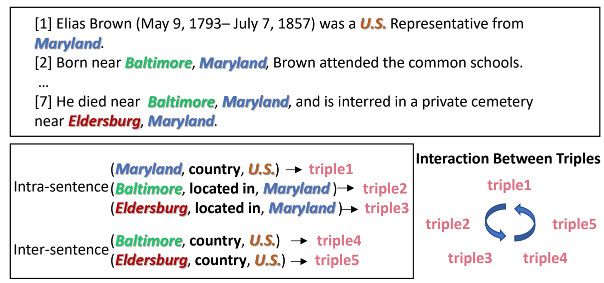

Figure 1: Example document with entity pairs and relations from DocRED. Entity mentions and relations only involved in these relation instances are colored.

to classify the relations of multiple entity pairs at once. Besides, the subject and object entities involved in a relation may appear in different sentences. Therefore a relation cannot be identified based solely on a single sentence. For example, as shown in Figure 1, it is easy to identify the intra-sentence relations, such as (*Maryland*, country, *U.S.*), (*Baltimore*, located_in, *Maryland*), and (*Eldersburg*, located_in, *Maryland*), owing to the occurrence of entities in the same sentence. However, it is more challenging for a model to recognize inter-sentence relations, such as those between *Eldersburg* and *U.S.* and between *Baltimore* and *U.S.* because these mentions occur in different sentences and have long-distance dependencies.

To extract relations among these inter-sentence entity pairs, most current studies constructed document-level graph module based on heuristics, structured attention or dependency structures [Peng *et al.*, 2017; Christopoulou *et al.*, 2019; Nan *et al.*, 2020; Zeng *et al.*, 2020; Wang *et al.*, 2020a], followed by reasoning with graph neural models. Meanwhile, considering the transformer architecture can implicitly model long-distance dependencies, some studies [Wang *et al.*, 2019; Tang *et al.*, 2020; Zhou *et al.*, 2021] directly applied pre-trained language models rather than explicit graph reasoning. In general, current approaches obtain entity representation via information passing through nodes on document-level graphs or transformer-based structure learning. However, they mainly focus on token-level syntactic features or contextual information rather than global interactions between entity

---
* Equal contribution and shared co-first authorship.
† Corresponding author.
[1]The code and datasets are available in https://github.com/zjunlp/DocuNet.

pairs, neglecting the interdependency among the multiple relations in one context.

Concretely, the interdependency among multiple triples is advantageous and can provide guidance for relation classification in the case of many entities. For example, if the intra-sentence relation (*Maryland*, country, *U.S.*) has been identified, it is implausible for *U.S.* to be in any other person-social relationship, such as "is the father of...". Besides, according to the triples that *Eldersburg* is located in *Maryland* and *Maryland* belongs to *U.S.*, we can infer that *Eldersburg* belongs to *U.S.*. As described above, each relation triple can provide information to other relation triples in the same text.

To capture the interdependency among the multiple triples, we reformulate the document-level RE task as an entity-level classification problem [Jiang *et al.*, 2019], also known as table filling [Miwa and Sasaki, 2014; Gupta *et al.*, 2016], as shown in Figure 2. It is analogous to semantic segmentation (a well-known computer vision task), whose goal is to label each pixel of the image with the corresponding represented class by convolution network. Inspired by the above, we propose a novel model called **Doc**ument **U**-shaped **N**etwork (**DocuNet**), which formulates document-level RE as semantic segmentation. In this manner, given relevant features between entity pairs as an image, the model predicts the relation type for each entity pair as a pixel-level mask. Specifically, we introduce an encoder module to capture the context information of entities and a U-shaped segmentation module over the image-style feature map to capture global interdependency among triples. We further propose a balanced softmax method to handle the imbalance relation distribution. Our contributions can be summarized as follows:

- To the best of our knowledge, this is the first approach that regards document-level RE as a semantic segmentation task.
- We introduce the model DocuNet to capture both local context information and global interdependency among triples for document-level RE.
- Experimental results on three benchmark datasets show that our model DocuNet can achieve state-of-the-art performance compared with baselines.

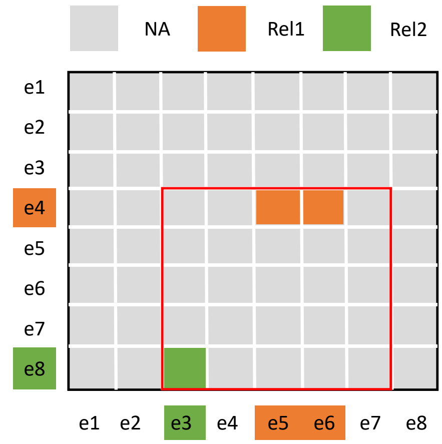

Figure 2: Illustration of the entity-level relation matrix applied in our formulation. Each cell belongs to one relation type.

2 Related Work

Previous relation extraction approaches mainly concentrate on identifying the relation between two entities within a sentence. Many approaches [Zeng *et al.*, 2015; Feng *et al.*, 2018; Zhang *et al.*, 2018; Zhang *et al.*, 2019; Zhang *et al.*, 2020b; Zhang *et al.*, 2020a; Wang *et al.*, 2020b; Ye *et al.*, 2021; Yu *et al.*, 2020; Wang *et al.*, 2020b; Wu *et al.*, 2021; Chen *et al.*, 2021; Zheng *et al.*, 2021] have been proposed to tackle the sentence-level RE task effectively. However, sentence-level RE faces an inevitable restriction in that many real-world relations can only be extracted by reading multiple sentences. For this reason, document-level RE appeals to many researchers [Tang *et al.*, 2020; Nan *et al.*, 2020; Zeng *et al.*, 2020; Wang *et al.*, 2020a; Xiao *et al.*, 2020].

Various approaches for document-level RE mainly include graph-based models and transformer-based models. Graph-based approaches are now widely adopted in RE because of their effectiveness and strength in relational reasoning. Jia *et al.* [2019] proposed a model that combines representations learned over various text spans throughout the document and across the sub-relation hierarchy. Christopoulou *et al.* [2019] proposed an edge-oriented graph neural model (EoG) for document-level RE. Li *et al.* [2020] characterized the complex interaction between sentences and potential relation instances with a graph-enhanced dual attention network (GEDA). Zhang *et al.* [2020c] proposed a novel graph-based model with a Dual-tier Heterogeneous Graph (DHG), which contains a structure modeling layer followed by a relation reasoning layer. Zhou *et al.* [2020] proposes a global context-enhanced graph convolutional network (GCGCN), composed of entities as nodes and the contexts of entity pairs as edges between nodes. Wang *et al.* [2020a] proposed a novel model (GLRE) that encodes the document information in terms of global and local entity representations as well as context relation representations. Nan *et al.* [2020] proposed a novel model (LSR) that enables relational reasoning across sentences by automatically inducing a latent document-level graph. Zeng *et al.* [2020] proposed the graph aggregation-and-inference network (GAIN) with double graphs for document-level RE. Xu *et al.* [2021] proposed an encoder-classifier reconstructor model (HeterGSAN), which manages to reconstruct the ground-truth path dependencies from the graph representation. Explicit graph reasoning can bridge the gap between entities that occur in different sentences, thus mitigating long-distance dependency and achieving promising performance.

In contrast, considering the transformer architecture can implicitly model long-distance dependencies, some researchers directly leverage pre-trained language models without generating document graphs. Wang *et al.* [2019] proposed a two-step training paradigm on DocRED using BERT as pre-trained word embedding. They observed an imbalance in the distribution of relation and disentangled the relation identification and classification for better inference. Tang *et al.* [2020] proposed a hierarchical inference network (HIN) to make full use of the abundant information from the entity,

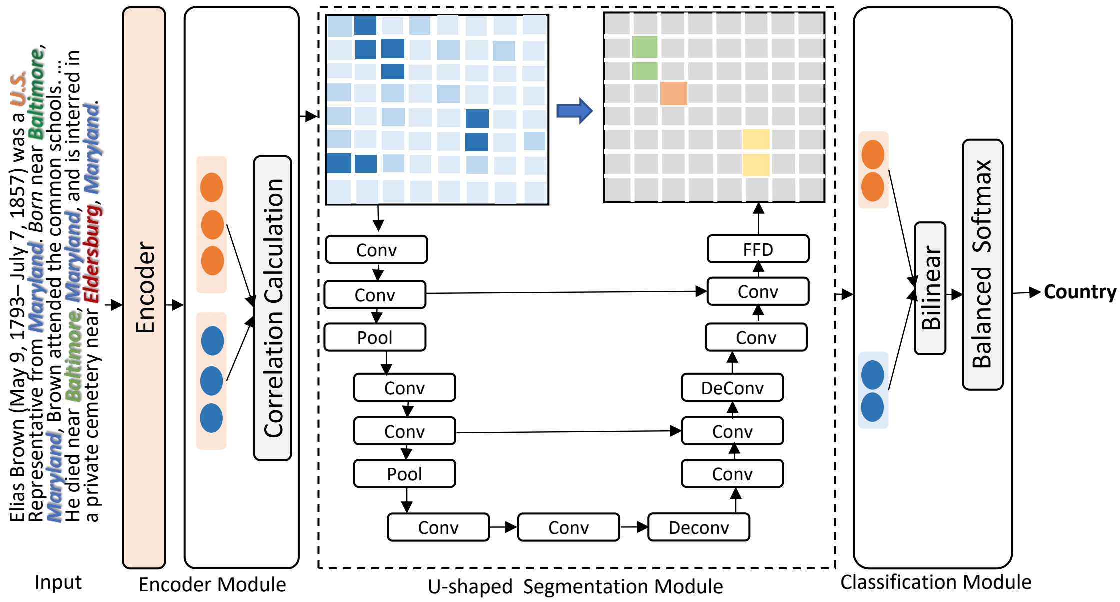

Figure 3: Architecture of our **Docu**ment **U**-shaped **N**etwork (**DocuNet**) (Best viewed in color).

sentence, and document levels to perform hierarchical reasoning. Zhou *et al.* [2021] proposed a novel transformer-based model (ATLOP) of adaptive thresholding and localized context pooling based on BERT. However, most previous studies focused on the local entity representation, regardless of the high-level global connections between triples, which overlooked the interdependency between multiple relations.

On the one hand, our work is inspired by [Jin *et al.*, 2020], which was the first to consider the issue of global interaction between relations, and there have been few studies on RE. On the other hand, as these studies[Nguyen and Grishman, 2015; Shen and Huang, 2016] have done, convolutional neural networks have been long used in the relation extraction area, which enlightens us to pay attention to the role of CNN in extracting information of the image-style feature map. Hence, our work is also related to the study of [Liu *et al.*, 2020], who formulated incomplete utterance rewriting as a semantic segmentation task and motivated us to study the RE problem from a computer vision perspective. In this study, we leveraged the U-Net [Ronneberger *et al.*, 2015], which consists of a contracting path to capture context and a symmetric expanding path that enables precise localization. To the best of our knowledge, this is the first approach to formulate RE as a semantic segmentation task.

3 Methodology

3.1 Preliminary

We first introduce the problem definition. With a document d containing a set of entities $\{e_i\}_{i=1}^{n}$, the task is to extract the relations between entity pairs (e_s, e_o). In one document, each entity e_i may occur multiple times. To model relation extraction between e_s and e_o, we define a $N \times N$ matrix Y, where entry $Y_{s,o}$ indicates the relation type between e_s and e_o. Then, we obtain the output of matrix Y, analogous to the task of semantic segmentation. Entities in Y are arranged according to their first appearance in the document. We obtain the feature map via the entity-to-entity relevance estimation and take the feature map as an image. Note that the output entity-level relation matrix Y is parallel to the pixel-level mask in semantic segmentation, which bridges relation extraction and semantic segmentation. Our approach can also be applied to sentence-level relation extraction. Since the document has relatively more entities, thus, entity-level relation matrix can learn more global information to boost the performance.

3.2 Encoder Module

Given the document $d = [x_t]_{t=1}^{L}$, we insert special symbols "$< e >$" and "$< /e >$" at the start and end of mentions to mark the entity positions. We leverage the pre-trained language model as an encoder to obtain the embedding as follows:

$$H = [h_1, h_2, ..., h_L] = \text{Encoder}([x_1, x_2, ..., x_L]). \quad (1)$$

where h_i is the embedding of the token x_i. Note that some documents are longer than 512, we thus leverage a *dynamic window* to encode whole documents. We average the embeddings of overlapping tokens of different windows to obtain the final representations. Then, we utilize the embeddings of "$< e >$" to represent mention following [Verga *et al.*, 2018]. We leverage a smooth version of max pooling, namely, log-sumexp pooling [Jia *et al.*, 2019] each entity e_i, to obtain the entity embedding $\mathbf{e}_i$:

$$\mathbf{e}_i = \log \sum_{j=1}^{N_{e_i}} \exp\left(\mathbf{m_j}\right). \quad (2)$$

This pooling accumulates signals from mentions in the document. Thus, we obtain the entity embedding $\mathbf{e}_i$.

We calculate the entity-level relation matrix based on entity-to-entity relevance. For each entity e_i in the matrix, their relevance is captured by a D-dimensional feature vector $\mathbf{F}(e_s, e_o)$. We introduce two strategies for computing $\mathbf{F}(e_s, e_o)$, namely, *similarity-based* method and *context-based* method. Similarity-based method is produced by concatenating operation result of element-wise similarity, cosine similarity and bi-linear similarity between e_s and e_o as:

$$\mathbf{F}(e_s, e_o) = \left[e_s \odot e_o; cos(e_s,e_o); e_s W_1 e_o\right], \tag{3}$$

For the context-based strategy, we leverage entity-aware attention with affine transformation to obtain the feature vector as follows:

$$\mathbf{F}(e_s, e_o) = W_2 H a^{(s,o)} \tag{4}$$

$$a^{(s,o)} = softmax(\sum_{i=1}^{K} A_i^s \cdot A_i^o) \tag{5}$$

where $a^{(s,o)}$ is the attention weight for entity-aware attention and A_i^s refers to the tokens' importance to the i-th entity, H is the document embedding, W_1, W_2 is the learnable weight matrix, K is the number of head in the transformer.

3.3 U-shaped Segmentation Module

Taking the entity-level relation matrix $\mathbf{F} \in R^{N \times N \times D}$ as a D-channel image, we formulate the document-level relation prediction as the pixel-level mask in F. where N is the largest number of entities, counted from all the dataset samples. Specifically N is the largest number of entities, counted from all the dataset samples. To this end, we utilize U-Net [Ronneberger *et al.*, 2015], which is a famous semantic segmentation model in computer vision. As can be seen in Figure 3, the module is formed as a U-shaped segmentation structure, which contains two down-sampling blocks and two up-sampling blocks with skip connections. On the one hand, each down-sampling block has two subsequent max pooling and separate convolution modules. Further, the number of channels is doubled in each down-sampling block. As it shows in the Figure 2, the segmentation area in the entity-level relation matrix refers to the co-occurrence of relations between entity pairs. The U-shaped segmentation structure can promote the information exchange between entity pairs in the receptive field analogy to implicit reasoning. Specifically, CNN and down-sampling block can enlarge the receptive field of current entity pair embedding $\mathbf{F}(e_s, e_o)$, thus, providing rich global information for representation learning. On the other hand, the model has two up-sampling blocks with a subsequent deconvolution neural network and two separate convolution modules. Different from down-sampling, the number of channels is halved in each up-sampling block, which can distribute the aggregated information to each pixel.

Finally, we incorporate an encoding module and a U-shaped segmentation module to capture both local and global information Y as follows:

$$\mathbf{Y} = U(W_3 \mathbf{F}) \tag{6}$$

where U and $\mathbf{Y} \in R^{N \times N \times D'}$ denote the U-shaped segmentation module and entity-level relation matrix respectively. W_3 is the learnable weight matrix in order to reduce the dimension of F and D' is much smaller than D.

Statistics / Dataset	DocRED	CDR	GDA
# Train	3,053	500	23,353
# Dev	1,000	500	5,839
# Test	1,000	500	1,000
# Relations	97	2	2
Avg. # entities per Doc.	19.5	7.6	5.4
Avg. # Ment. per Ent.	1.4	2.7	3.3

Table 1: Statistics of the experimental datasets.

3.4 Classification Module

Given the entity pair embedding $\mathbf{e}_s$ and $\mathbf{e}_o$ with the entity-level relation matrix Y, we map them to hidden representations z with a feedforward neural network. Then, we obtain the probability of relation via a bilinear function. Formally, we have:

$$z_s = \tanh\left(W_s \mathbf{e_s} + Y_{s,o}\right), \tag{7}$$

$$z_o = \tanh\left(W_o \mathbf{e_o} + Y_{s,o}\right), \tag{8}$$

$$\mathrm{P}\left(r|e_s, e_o\right) = \sigma\left(z_s W_r z_o + b_r\right), \tag{9}$$

where $Y_{s,o}$ is the entity-pair representation of (s, o) in matrix Y, $W_r \in R^{d \times d}$, $b_r \in R$, $W_s \in R^{d \times d}$, and $W_o \in R^{d \times d}$, are learnable parameters.

Since previous work [Wang *et al.*, 2019] observed that there is an imbalance relation distribution for RE (many entity pairs have relation of *NA*), we introduce a balanced softmax method for training, which is inspired by the circle loss [Sun *et al.*, 2020] from computer vision. Specifically, we introduce an additional category 0, hoping that the scores of the target category are all greater than s_0 and the scores of the non-target categories are all less than s_0. Formally, we have:

$$L = \log\left(e^{s_0} + \sum_{i \in \Omega_{neg}} e^{s_i}\right) + \log\left(e^{-s_0} + \sum_{j \in \Omega_{pos}} e^{-s_j}\right). \tag{10}$$

For simplicity, we set the threshold as zero and have the following:

$$L = \log\left(1 + \sum_{i \in \Omega_{neg}} e^{s_i}\right) + \log\left(1 + \sum_{j \in \Omega_{pos}} e^{-s_j}\right). \tag{11}$$

4 Experiments

4.1 Dataset

We evaluated our DocuNet model on three document-level RE datasets. We listed the dataset statistics in Table 1.

- **DocRED** [Yao *et al.*, 2019] is a large-scale document-level relation extraction dataset by crowdsourcing. DocRED contains 3,053/1,000/1,000 instances for training, validating and test, respectively.

Model	Dev		Test	
	Ign F_1	F_1	Ign F_1	F_1
GEDA-BERT$_{\text{base}}$ [Li *et al.*, 2020]	54.52	56.16	53.71	55.74
LSR-BERT$_{\text{base}}$ [Nan *et al.*, 2020]	52.43	59.00	56.97	59.05
GLRE-BERT$_{\text{base}}$ [Wang *et al.*, 2020a]	-	-	55.40	57.40
GAIN-BERT$_{\text{BASE}}$ [Zeng *et al.*, 2020]	59.14	61.22	59.00	61.24
HeterGSAN-BERT$_{\text{base}}$ [Xu *et al.*, 2021]	58.13	60.18	57.12	59.45
BERT$_{\text{base}}$ [Wang *et al.*, 2019]	-	54.16	-	53.20
BERT-TS$_{\text{base}}$ [Wang *et al.*, 2019]	-	54.42	-	53.92
HIN-BERT$_{\text{base}}$ [Tang *et al.*, 2020]	54.29	56.31	53.70	55.60
CorefBERT$_{\text{base}}$ [Ye *et al.*, 2020]	55.32	57.51	54.54	56.96
ATLOP-BERT$_{\text{base}}$ [Zhou *et al.*, 2021]	59.22	61.09	59.31	61.30
DocuNet-BERT$_{\text{base}}$	**59.86±0.13**	**61.83±0.19**	**59.93**	**61.86**
BERT$_{\text{large}}$ [Ye *et al.*, 2020]	56.67	58.83	56.47	58.69
CorefBERT$_{\text{large}}$ [Ye *et al.*, 2020]	56.82	59.01	56.40	58.83
RoBERTa$_{\text{large}}$ [Ye *et al.*, 2020]	57.14	59.22	57.51	59.62
CorefRoBERTa$_{\text{large}}$ [Ye *et al.*, 2020]	57.35	59.43	57.90	60.25
ATLOP-RoBERTa$_{\text{large}}$ [Zhou *et al.*, 2021]	61.32	63.18	61.39	63.40
DocuNet-RoBERTa$_{\text{large}}$	**62.23±0.12**	**64.12±0.14**	**62.39**	**64.55**

Table 2: Results (%) on the development and test set of DocRED. We run experiments five times with different random seeds and report the mean and standard deviation on the development set. We report the official test score on the CodaLab scoreboard with the best checkpoint on the development set.

Model	CDR	GDA
BRAN [Verga *et al.*, 2018]	62.1	-
EoG [Christopoulou *et al.*, 2019]	63.6	81.5
LSR [Nan *et al.*, 2020]	64.8	82.2
DHG [Zhang *et al.*, 2020c]	65.9	83.1
GLRE [Wang *et al.*, 2020a]	68.5	-
SciBERT$_{\text{base}}$ [Beltagy *et al.*, 2019]	65.1	82.5
ATLOP-SciBERT$_{\text{base}}$ [Zhou *et al.*, 2021]	69.4	83.9
DocuNet-SciBERT$_{\text{base}}$	**76.3±0.40**	**85.3±0.50**

Table 3: Results (%) on the biomedical datasets CDR and GDA.

- **CDR** [Li *et al.*, 2016] is a relation extraction dataset in the biomedical domain, which is aimed to infer the interactions between chemical and disease concepts.
- **GDA** [Wu *et al.*, 2019] is a dataset in the biomedical domain, which consists of 23,353 training samples. Differently, the dataset is aimed to predict the interactions between disease concepts and genes.

4.2 Experimental Settings

Our model was implemented based on Pytorch. We used cased BERT-base, or RoBERTa-large as the encoder on DocRED and SciBERT-base [Beltagy *et al.*, 2019] on CDR and GDA. We optimize our model with AdamW using learning rates $2e{-}5$ with a linear warmup for the first 6% of steps. We set the matrix size $N = 42$. The context-based strategy is utilized by default. We tuned the hyperparameters on the development set. We trained on one NVIDIA V100 16GB GPU and evaluated our model with Ign F1, and F1 following [Yao *et al.*, 2019].

4.3 Results on the DocRED Dataset

We compare DocuNet with graph-based models, including GEDA [Li *et al.*, 2020], LSR [Nan *et al.*, 2020], GLRE [Wang *et al.*, 2020a] and GAIN [Zeng *et al.*, 2020], HeterGSAN [Xu *et al.*, 2021]; and transformer-based models, including BERT$_{\text{base}}$ [Wang *et al.*, 2019], BERT-TS$_{\text{base}}$ [Wang *et al.*, 2019], HIN-BERT$_{\text{base}}$ [Tang *et al.*, 2020], CorefBERT$_{\text{base}}$ [Ye *et al.*, 2020], and ATLOP$_{\text{base}}$ on the DocRED dataset. From the Table 2, we observed that our approach DocuNet-BERT$_{\text{base}}$ obtains better results than ATLOP-BERT$_{\text{base}}$. Moreover, we found that our DocuNet model obtain a new state-of-the-art result with with RoBERTa-large. As of the IJCAI deadline on 20th of January 2021, **we held the first position on the CodaLab scoreboard[2] under the alias *DocuNet* without external data**[3].

4.4 Results on the Biomedical Datasets

In the biomedical datasets, we compare DocuNet with lots of baselines including: BRAN [Verga *et al.*, 2018], EoG [Christopoulou *et al.*, 2019], LSR [Nan *et al.*, 2020], DHG [Zhang *et al.*, 2020c], GLRE [Wang *et al.*, 2020a] and ATLOP [Zhou *et al.*, 2021]. Following ATLOP[Zhou *et al.*, 2021], we utilize the SciBERT [Beltagy *et al.*, 2019] which

[2]https://competitions.codalab.org/competitions/20717#results

[3]The SSAN_ADAPT model leverages pre-training with external distance supervised data.

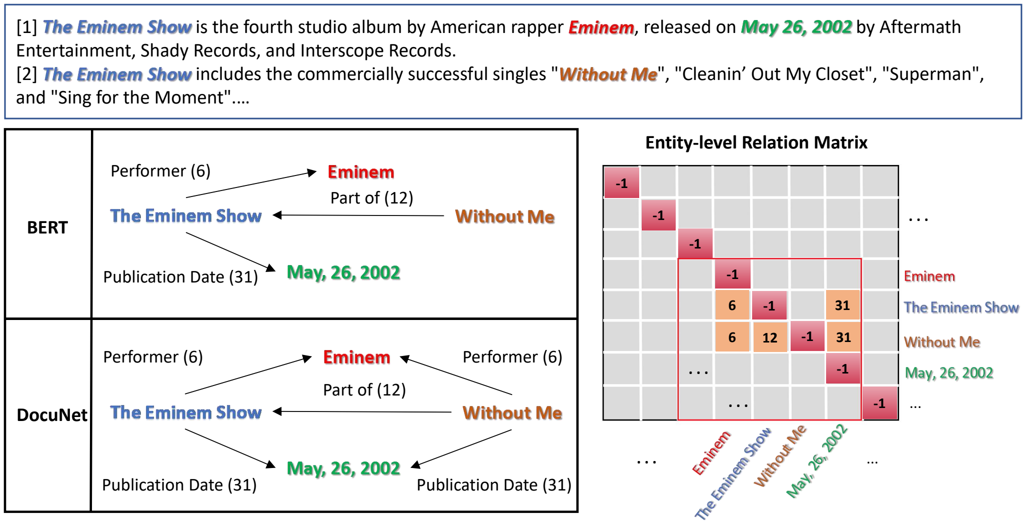

Figure 4: Case study on our proposed DocuNet and baseline model. The specific number in the figure indicates the corresponding label id.

Model	Ign F_1	F_1
DocuNet (Context-based)	**59.86**	**61.83**
DocuNet (Similarity-based)	59.04	60.92
w/o Balanced Softmax	58.56	60.51
w/o U-shaped Segmentation	57.51	59.65

Table 4: Ablation study of DocuNet on DocRED.

is pre-trained on the scientific publication corpora. From the Table 3, we observe model DocuNet-SciBERT$_{\text{base}}$ improved the F1 score by **6.9%** and **1.4%** on CDR and GDA compared with ATLOP-SciBERT$_{\text{base}}$,

4.5 Ablation Study

We conducted an ablation study experiment to validate the effectiveness of different components of our approach. **DocuNet (Similarity-based)** means directly using similarity functions strategy to calculate the correlation between two entities as the input matrix, rather than context-based strategy. **w/o U-shaped Segmentation** means that our segmentation module is replaced by a feed-forward neural network. **w/o balanced softmax** refers to the model only with binary cross-entropy loss. From Table 4, we observe that all models have a performance decay without each module, which indicates that both components are beneficial. Besides, we observed that the U-shaped segmentation module and balanced softmax module are most important to model performance and sensitive to F_1, leading to a drop of 2.18% and 1.32% in dev F_1 score respectively when removed from DocuNet. That reveals that global interdependency among triples captured by our model is effective for document-level RE. Moreover, compared with context-based strategy, our approach based on similarity functions strategy drop by 0.84 F_1, which illustrates the context-based strategy is advantageous.

4.6 Case Study

We follow GAIN [Zeng *et al.*, 2020] to select the same example and conduct a case study to further illustrate the effectiveness of our model DocuNet compared with the baseline. As shown in Figure 4, we notice that both BERT$_{base}$ and DocuNet-BERT$_{base}$ can successfully extract the "part of" relation between *"Without Me"* and *"The Eminem Show"*. However, only our model DocuNet-BERT$_{base}$ is able to deduce that the "performer" and "publication date" of *"Without Me"* are the same as those of *"The Eminem Show"*, namely, *"Eminem"* and *"May 26, 2002"*, respectively.

Intuitively we can observe that relation extraction mentioned above among those entities requires logical inference across sentences. This interesting observation indicates that our U-shaped segmentation structure over the entity-level relation matrix may implicitly conduct relational reasoning among entities.

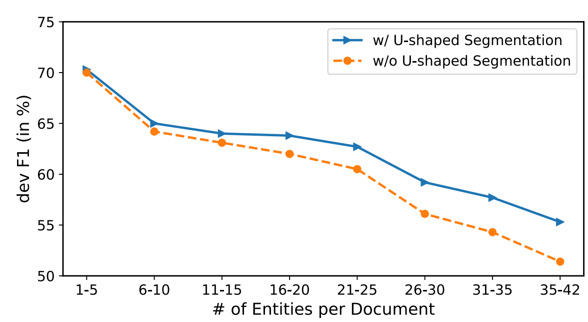

Figure 5: Dev results in terms of number of entities on DocRED.

4.7 Analysis

To assess the effectiveness of DocuNet in modeling global information for multiple entities, we evaluated models respectively trained with or without U-shaped segmentation module on different groups of development set in DocRED, which are divided by the number of entities. From Figure 5, we observe that the model w/ U-shaped segmentation module consistently outperforms the model w/o U-shaped segmentation module. We notice that when the number of entities increases, the improvement becomes larger. This indicates that our U-shaped segmentation module can implicitly learn the interdependency among the multiple triples in one context, thus improving the document-level RE performance.

5 Conclusion and Future Work

In this study, we took the first step in formulating document-level RE as a semantic segmentation task and introducing the Document U-shaped Network. Experimental results showed that our model could achieve better performance by capturing local and global information than baselines. We also empirically observe that convolution over entity-entity relation matrix may implicitly conduct relational reasoning among entities. In the future, we plan to apply our approach to other span-level classification tasks, such as aspect-based sentiment analysis and nest named recognition.

Acknowledgments

We want to express gratitude to the anonymous reviewers for their hard work and kind comments. We thank Ning Ding for helpful discussions and feedback on this paper. This work is funded by National Key R&D Program of China (Funding No. 2018YFB1402800), NSFC91846204.

References

[Beltagy *et al.*, 2019] Iz Beltagy, Kyle Lo, and Arman Cohan. Scibert: A pretrained language model for scientific text. In *EMNLP/IJCNLP*, 2019.

[Chen *et al.*, 2021] Xiang Chen, Xin Xie, Ningyu Zhang, Jiahuan Yan, Shumin Deng, Chuanqi Tan, Fei Huang, Luo Si, and Huajun Chen. Adaprompt: Adaptive prompt-based finetuning for relation extraction. *CoRR*, abs/2104.07650, 2021.

[Christopoulou *et al.*, 2019] Fenia Christopoulou, Makoto Miwa, and Sophia Ananiadou. Connecting the dots: Document-level neural relation extraction with edge-oriented graphs. In *EMNLP/IJCNLP*, 2019.

[Feng *et al.*, 2018] Jun Feng, Minlie Huang, Li Zhao, Yang Yang, and Xiaoyan Zhu. Reinforcement learning for relation classification from noisy data. In *AAAI*, pages 5779–5786, 2018.

[Gupta *et al.*, 2016] Pankaj Gupta, Hinrich Schütze, and Bernt Andrassy. Table filling multi-task recurrent neural network for joint entity and relation extraction. In *COLING*, pages 2537–2547, 2016.

[Jia *et al.*, 2019] Robin Jia, Cliff Wong, and Hoifung Poon. Document-level n-ary relation extraction with multiscale representation learning. In *NAACL-HLT*, 2019.

[Jiang *et al.*, 2019] Zhengbao Jiang, Wei Xu, Jun Araki, and Graham Neubig. Generalizing natural language analysis through span-relation representations. In *ACL*, 2019.

[Jin *et al.*, 2020] Zhijing Jin, Yongyi Yang, Xipeng Qiu, and Zheng Zhang. Relation of the relations: A new paradigm of the relation extraction problem. *arXiv preprint arXiv:2006.03719*, 2020.

[Li *et al.*, 2016] J. Li, Yueping Sun, Robin J. Johnson, Daniela Sciaky, Chih-Hsuan Wei, Robert Leaman, A. P. Davis, C. Mattingly, Thomas C. Wiegers, and Zhiyong Lu. Biocreative v cdr task corpus: a resource for chemical disease relation extraction. *Database: The Journal of Biological Databases and Curation*, 2016, 2016.

[Li *et al.*, 2020] Bo Li, Wei Ye, Zhonghao Sheng, Rui Xie, Xiangyu Xi, and Shikun Zhang. Graph enhanced dual attention network for document-level relation extraction. In *COLING*, pages 1551–1560, 2020.

[Liu *et al.*, 2020] Qian Liu, Bei Chen, Jian-Guang Lou, Bin Zhou, and Dongmei Zhang. Incomplete utterance rewriting as semantic segmentation. In *EMNLP*, 2020.

[Miwa and Sasaki, 2014] Makoto Miwa and Yutaka Sasaki. Modeling joint entity and relation extraction with table representation. In *EMNLP*, 2014.

[Nan *et al.*, 2020] G. Nan, Zhijiang Guo, Ivan Sekulic, and W. Lu. Reasoning with latent structure refinement for document-level relation extraction. In *ACL*, 2020.

[Nguyen and Grishman, 2015] Thien Huu Nguyen and Ralph Grishman. Relation extraction: Perspective from convolutional neural networks. In Phil Blunsom, Shay B. Cohen, Paramveer S. Dhillon, and Percy Liang, editors, *Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing, VS@NAACL-HLT 2015, June 5, 2015, Denver, Colorado, USA*, pages 39–48. The Association for Computational Linguistics, 2015.

[Peng *et al.*, 2017] Nanyun Peng, Hoifung Poon, Chris Quirk, Kristina Toutanova, and Wen tau Yih. Cross-sentence n-ary relation extraction with graph lstms. *TACL*, 5:101–115, 2017.

[Ronneberger *et al.*, 2015] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, volume 9351 of *LNCS*, pages 234–241. Springer, 2015.

[Shen and Huang, 2016] Yatian Shen and Xuanjing Huang. Attention-based convolutional neural network for semantic relation extraction. In Nicoletta Calzolari, Yuji Matsumoto, and Rashmi Prasad, editors, *COLING 2016, 26th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, December 11-16, 2016, Osaka, Japan*, pages 2526–2536. ACL, 2016.

[Sun *et al.*, 2020] Yifan Sun, Changmao Cheng, Yuhan Zhang, Chi Zhang, Liang Zheng, Zhongdao Wang, and Yichen Wei. Circle loss: A unified perspective of pair similarity optimization. In *CVPR*, pages 6398–6407, 2020.

[Tang *et al.*, 2020] Hengzhu Tang, Yanan Cao, Zhenyu Zhang, Jiangxia Cao, Fang Fang, Shi Wang, and Pengfei Yin. Hin: Hierarchical inference network for document-level relation extraction. In *PAKDD*, 2020.

[Verga *et al.*, 2018] Pat Verga, Emma Strubell, and Andrew McCallum. Simultaneously self-attending to all mentions for full-abstract biological relation extraction. In *NAACL-HLT*, 2018.

[Wang *et al.*, 2019] Hong Wang, Christfried Focke, Rob Sylvester, Nilesh Mishra, and William W. J. Wang. Fine-tune bert for docred with two-step process. *ArXiv*, abs/1909.11898, 2019.

[Wang *et al.*, 2020a] Difeng Wang, Wei Hu, Ermei Cao, and Weijian Sun. Global-to-local neural networks for document-level relation extraction. In *EMNLP*, pages 3711–3721, 2020.

[Wang *et al.*, 2020b] Zifeng Wang, Rui Wen, Xi Chen, Shao-Lun Huang, Ningyu Zhang, and Yefeng Zheng. Finding influential instances for distantly supervised relation extraction. *CoRR*, abs/2009.09841, 2020.

[Wu *et al.*, 2019] Y. Wu, Ruibang Luo, H. Leung, H. Ting, and T. Lam. Renet: A deep learning approach for extracting gene-disease associations from literature. In *RECOMB*, 2019.

[Wu *et al.*, 2021] Tongtong Wu, Xuekai Li, Yuan-Fang Li, Reza Haffari, Guilin Qi, Yujin Zhu, and Guoqiang Xu. Curriculum-meta learning for order-robust continual relation extraction. In *AAAI*, 2021.

[Xiao *et al.*, 2020] Chaojun Xiao, Yuan Yao, Ruobing Xie, Xu Han, Zhiyuan Liu, Maosong Sun, Fen Lin, and Leyu Lin. Denoising relation extraction from document-level distant supervision. In *EMNLP*, 2020.

[Xu *et al.*, 2021] Wang Xu, Kehai Chen, and Tiejun Zhao. Document-level relation extraction with reconstruction. In *AAAI*, 2021.

[Yao *et al.*, 2019] Yuan Yao, D. Ye, Peng Li, Xu Han, Yankai Lin, Zhenghao Liu, Z. Liu, Lixin Huang, Jie Zhou, and M. Sun. Docred: A large-scale document-level relation extraction dataset. In *ACL*, 2019.

[Ye *et al.*, 2020] Deming Ye, Yankai Lin, Jiaju Du, Zhenghao Liu, Maosong Sun, and Zhiyuan Liu. Coreferential reasoning learning for language representation. In *EMNLP*, 2020.

[Ye *et al.*, 2021] Hongbin Ye, Ningyu Zhang, Shumin Deng, Mosha Chen, Chuanqi Tan, Fei Huang, and Huajun Chen. Contrastive triple extraction with generative transformer. In *AAAI*, 2021.

[Yu *et al.*, 2020] Haiyang Yu, Ningyu Zhang, Shumin Deng, Hongbin Ye, Wei Zhang, and Huajun Chen. Bridging text and knowledge with multi-prototype embedding for few-shot relational triple extraction. In *COLING*, 2020.

[Zeng *et al.*, 2015] Daojian Zeng, Kang Liu, Yubo Chen, and Jun Zhao. Distant supervision for relation extraction via piecewise convolutional neural networks. In *EMNLP*, pages 1753–1762, 2015.

[Zeng *et al.*, 2020] Shuang Zeng, Runxin Xu, Baobao Chang, and Lei Li. Double graph based reasoning for document-level relation extraction. In *EMNLP*, 2020.

[Zhang *et al.*, 2018] Ningyu Zhang, Shumin Deng, Zhanlin Sun, Xi Chen, Wei Zhang, and Huajun Chen. Attention-based capsule networks with dynamic routing for relation extraction. In *EMNLP*, 2018.

[Zhang *et al.*, 2019] Ningyu Zhang, Shumin Deng, Zhanlin Sun, Guanying Wang, Xi Chen, Wei Zhang, and Huajun Chen. Long-tail relation extraction via knowledge graph embeddings and graph convolution networks. In *NAACL-HLT*, 2019.

[Zhang *et al.*, 2020a] Ningyu Zhang, Shumin Deng, Zhen Bi, Haiyang Yu, Jiacheng Yang, Mosha Chen, Fei Huang, Wei Zhang, and Huajun Chen. Openue: An open toolkit of universal extraction from text. In Qun Liu and David Schlangen, editors, *EMNLP (Demo)*, pages 1–8. Association for Computational Linguistics, 2020.

[Zhang *et al.*, 2020b] Ningyu Zhang, Shumin Deng, Zhanlin Sun, Jiaoyan Chen, Wei Zhang, and Huajun Chen. Relation adversarial network for low resource knowledge graph completion. In *Proceedings of The Web Conference 2020*, 2020.

[Zhang *et al.*, 2020c] Zhenyu Zhang, Bowen Yu, Xiaobo Shu, Tingwen Liu, Hengzhu Tang, Wang Yubin, and Li Guo. Document-level relation extraction with dual-tier heterogeneous graph. In *COLING*, pages 1630–1641, 2020.

[Zhang *et al.*, 2021a] Ningyu Zhang, Qianghuai Jia, Shumin Deng, Xiang Chen, Hongbin Ye, Hui Chen, Huaixiao Tou, Gang Huang, Zhao Wang, Nengwei Hua, and Huajun Chen. Alicg: Fine-grained and evolvable conceptual graph construction for semantic search at alibaba. In *KDD*, 2021.

[Zhang *et al.*, 2021b] Shengyu Zhang, Dong Yao, Zhou Zhao, Tat-Seng Chua, and Fei Wu. Causerec: Counterfactual user sequence synthesis for sequential recommendation. In *SIGIR*, 2021.

[Zheng *et al.*, 2021] Hengyi Zheng, Rui Wen, Xi Chen, Yifan Yang, Yunnan Zhang, Ziheng Zhang, Ningyu Zhang, Bin Qin, Xu Ming, and Yefeng Zheng. Prgc: Potential relation and global correpondence based joint relational triple extraction. In *ACL*, 2021.

[Zhou *et al.*, 2020] Huiwei Zhou, Yibin Xu, Weihong Yao, Zhe Liu, Chengkun Lang, and Haibin Jiang. Global context-enhanced graph convolutional networks for document-level relation extraction. In *COLING*, pages 5259–5270, 2020.

[Zhou *et al.*, 2021] Wenxuan Zhou, Kevin Huang, Tengyu Ma, and Jing Huang. Document-level relation extraction with adaptive thresholding and localized context pooling. In *AAAI*, 2021.