

Discriminative Reasoning for Document-level Relation Extraction

Wang Xu¹, Kehai Chen² and Tiejun Zhao¹

¹Harbin Institute of Technology, Harbin, China

²National Institute of Information and Communications Technology, Kyoto, Japan

xuwang@hit-mtlab.net, khchen@nict.go.jp, tjzhao@hit.edu.cn

Abstract

Document-level relation extraction (DocRE) models generally use graph networks to implicitly model the reasoning skill (i.e., pattern recognition, logical reasoning, coreference reasoning, etc.) related to the relation between one entity pair in a document. In this paper, we propose a novel discriminative reasoning framework to explicitly model the paths of these reasoning skills between each entity pair in this document. Thus, a discriminative reasoning network is designed to estimate the relation probability distribution of different reasoning paths based on the constructed graph and vectorized document contexts for each entity pair, thereby recognizing their relation. Experimental results show that our method outperforms the previous state-of-the-art performance on the large-scale DocRE dataset. The code is publicly available at <https://github.com/xwjim/DRN>.

1 Introduction

Document-level relation extraction (DocRE) aims to extract relations among entities within a document which requires multiple reasoning skills (i.e., pattern recognition, logical reasoning, coreference reasoning, and common-sense reasoning) (Yao et al., 2019). Generally, the input document is constructed as a structural graph-based on syntactic trees, coreference or heuristics to represent relation information between all entity pairs (Nan et al., 2020; Zeng et al., 2020; Xu et al., 2021). Thus, graph neural networks are applied to the constructed structural graph to model these reasoning skills. After performing multi-hop graph convolution, the feature representations of two entities are concatenated to recognize their relation by the classifier, achieving state-of-the-art performance in the DocRE task (Zeng et al., 2020; Xu et al., 2021). However, it is yet to be seen

Pattern Recognition: [1] <i>Me Musical Nephews</i> is a 1942 one-reel animated cartoon directed by Seymour Kenitel.	Supporting Evidence: 1
Relation: publication_data	
Common-sense Reasoning: [1] <i>William</i> and <i>Adelaide</i> had four children.	Supporting Evidence: 1
Relation: spouse	
Logical Reasoning: [1] Elias Brown (May 9, 1793– July 7, 1857) was a <i>U.S.</i> Representative from Maryland. [2] Born near <i>Baltimore</i> , Maryland, Brown attended the common schools. ... [7] He died near <i>Baltimore</i> , Maryland, and is interred in a private cemetery near Eldersburg, Maryland.	Supporting Evidence: 1,7
Relation: Country	
Coreference Reasoning: [1] <i>Dwight Tillery</i> is an American politician of the Democratic Party who is active in local politics of Cincinnati, Ohio. ... [3] He also holds a law degree from the <i>University of Michigan Law School</i> .	Supporting Evidence: 1,3
Relation: educated_at	

Figure 1: An example of different reasoning types. Different reasoning types have different reasoning processing.

whether modeling these reasoning skills implicitly is competitive with the intuitive reasoning skills between one entity pair in this document.

Figure 1 shows four kinds of reasoning skills for entity pairs in the DocRE dataset (Yao et al., 2019). First, take two entity pairs {"*Me Musical Nephews*", "1942"} and {"*William*", "*Adelaide*"} as examples, the intra-sentence reasoning concerns about the mentions inside the sentence, for example, "*Me Musical Nephews*" and "1942" for pattern recognition, and "*William*" and "*Adelaide*" for the common-sense reasoning. Also, the logical reasoning for entity pair {"*U.S.*", "*Baltimore*"} requires the reason path from "*U.S.*"→"*Maryland*" (bridge entity)→"*Baltimore*" while the coreference reasoning for entity pair {"*Dwight Tillery*", "*University of Michigan Law School*"} pays attention to the reason path from "*Dwight Tillery*"→"*He*" (reference word)→"*University of Michigan Law*

School". However, the advanced DocRE models generally use the universal multi-hop convolution networks to model these reasoning skills implicitly and do not consider the above intuitive reasoning skills explicitly, which may hinder the further improvement of DocRE.

To this end, we propose a novel discriminative reasoning framework to explicitly model the reasoning processing of these reasoning skills, such as intra-sentence reasoning (including pattern recognition and common-sense reasoning), logical reasoning, and coreference reasoning. Specifically, inspired by Xu et al.'s meta-path strategy, we extract the reasoning paths of the three reasoning skills discriminatively from the input document. Thus, a discriminative reasoning network is designed to estimate the relation probability distribution of different reasoning paths based on the constructed graph and vectorized document contexts for each entity pair, thereby recognizing their relation. In particular, there are the probabilities of multiple reasoning skills for each candidate relation between one entity pair, to ensure that all potential reasoning skills can be considered in the inference. In summary, our main contributions are as follows:

- We propose a discriminative reasoning framework to model the reasoning skills between two entities in a document. To the best of our knowledge, this is the first work to model different reasoning skills explicitly for enhancing the DocRE.
- Also, we introduce a discriminative reasoning network to encode the reasoning paths based on the constructed heterogeneous graph and the vectorized original document, thereby recognizing the relation between two entities by the classifier.
- Experimental results on the large-scale DocRE dataset show the effectiveness of the proposed method, especially outperform the recent state-of-the-art DocRE model.

2 Discriminative Reasoning Framework

In this section, we propose a novel discriminative reasoning framework to model different reasoning skills explicitly to recognize the relation between each entity pair in the input document. The discriminative reasoning framework contains three

parts: definition of reasoning paths, modeling reasoning discriminatively, and multi-reasoning based relation classification.

2.1 Definition of Reasoning Path

Formally, given one unstructured document comprised of N sentences $D=\{s_1, s_2, \dots, s_N\}$, each sentence is a sequence of words $s_n = \{s_n^1, s_n^2, \dots, s_n^J\}$ with the length $J_n=|s_n|$. The annotations include concept-level entities $\varepsilon = \{e_i\}_{i=1}^P$ as well as multiple occurrences of each entity under the same phrase or alias $e_i = \{m_i^{s_k}\}_{k=1}^Q$ ($m_i^{s_k}$ denotes the mention of e_i which occur in the sentence s_k) and their entity type (i.e. locations, organizations, and persons). The DocRE aims to extract the relation between two entities in ε , namely $P(r|e_i, e_j, D)$. For the simplification of reason skills, we first combine both pattern recognition and common-sense reasoning as the intra-sentence reasoning because they generally perform reasoning inside the sentence. Consequently, the original four kinds of the reasoning skills (Yao et al., 2019) are further refined as three reasoning skills: intra-sentence reasoning, logical reasoning, and coreference reasoning. Inspired by Xu et al.'s work, we also use the meta-path strategy to extract reasoning path for each reason skill, thereby representing the above three reasoning skills explicitly. Specifically, meta-paths for different reasoning skills are defined as follows:

- 1) **Intra-sentence reasoning path:** It is formally denoted as $PI_{ij}=m_i^{s_1} \circ s_1 \circ m_j^{s_1}$ for one entity pair $\{e_i, e_j\}$ inside the same sentence s_1 in the input document D . $m_i^{s_1}$ and $m_j^{s_1}$ are mentions related to two entities, respectively. "o" denotes one reasoning step on the reasoning path from e_i to e_j .
- 2) **Logical reasoning path:** The relation between one entity pair $\{e_i, e_j\}$ from sentences s_1 and s_2 is indirectly established by the occurrence bridge entity e_l for the logical reasoning. The reasoning path can be formally as $PL_{ij}=m_i^{s_1} \circ s_1 \circ m_l^{s_1} \circ m_l^{s_2} \circ s_2 \circ m_j^{s_2}$.
- 3) **Coreference reasoning path:** A reference word refers to one of two entities e_i and e_j , which occur in the same sentence as the other entity. We simplify the condition and assume that there is a coreference reasoning path

when the entities occur in different sentences. The reasoning path can be formally as $PC=m_i^{s_1} \circ s_1 \circ s_2 \circ m_j^{s_2}$.

Note that there are no entities in the defined reasoning path compare to the meta-path defined in Xu et al.’s work. This difference is mainly due to the following considerations: i) the reason path pays more attention to the mentions and referred sentences; ii) entities generally are contained by mentions; iii) it makes modeling of path reasoning more simple.

2.2 Modeling Reasoning Discriminatively

Based on the defined reasoning paths, we decompose the DocRE problem into three reasoning sub-tasks: intra-sentence reasoning (IR), logical reasoning (LR), and coreference reasoning (CR). Next, we introduce modeling of three sub-tasks in detail:

Modeling Intra-Sentence Reasoning. Given one entity pair $\{e_i, e_j\}$ and its reasoning path PI_{ij} in the sentence s_1 , the intra-sentence reasoning is modeled to recognize the relation between this entity pair based as follows:

$$R_{PI}(r) = P(r|e_i, e_j, PI_{ij}, D). \quad (1)$$

Modeling Logical Reasoning. Given one entity pair $\{e_i, e_j\}$ and its reasoning path PL_{ij} , the logical reasoning is modeled to recognize the relation between this entity pair based as follows:

$$R_{PL}(r) = P(r|e_i, e_j, PL_{ij}, D). \quad (2)$$

Since the e_l co-occur with the entity pair e_i and e_j respectively, the logical reasoning is further formally as follows:

$$R_{PL}(r) = P(r|e_i, e_j, e_l, PI_{il} \circ PI_{lj}, D). \quad (3)$$

where \circ denotes the connection of the paths.

Modeling Coreference Reasoning. Similarity, given one entity pair $\{e_i, e_j\}$ and its reasoning path PC_{ij} , the coreference reasoning is modeled to recognize the relation between this entity pair based as follows:

$$R_{PC}(r) = P(r|e_i, e_j, PC_{ij}, D). \quad (4)$$

2.3 Multi-reasoning Based Relation Classification

In the DocRE task, one entity usually involves multiple relationships which rely on different

reasoning types. Thus, the relation between one entity pair may be reasoned by multiple types of reasoning rather than one single reasoning type. Based on the proposed three reasoning sub-tasks, the relation reasoning between one entity pair is regarded as a multi-reasoning classification problem. Formally, we select the reasoning type with max probability to recognize the relation between each entity pair as follows:

$$P(r|e_i, e_j, D) = \max[R_{PI}(r), R_{PL}(r), R_{PC}(r)]. \quad (5)$$

In addition, there are often multiple reason paths between two entities for one reasoning type. Thus, the classification probability in Eq.(5) can be rewritten as follows:

$$P(r|e_i, e_j, D) = \max[\{R_{PI_1}(r), \dots, R_{PI_K}(r)\}, \{R_{PL_1}(r), \dots, R_{PL_K}(r)\}, \{R_{PC_1}(r), \dots, R_{PC_K}(r)\}], \quad (6)$$

where K is the number of reasoning paths for one reasoning skill, which is the same to each reasoning skill for simplicity. Note that all the entity pairs have at least one reasoning path from one of three defined reasoning sub-tasks. When the number of reasoning paths is greater than K for one reasoning sub-task, we choose the K first reasoning paths, otherwise we use the actual reasoning paths.

3 Discriminative Reasoning Network

In this section, we design a discriminative reasoning network (DRN) to model three defined reasoning sub-tasks for recognizing the relation between two entities in a document. Follow Zeng et al. and Zhou et al.’s work, we use two kinds of context representations (heterogeneous graph context representation and document-level context representation) to model different reasoning paths discriminatively in Eq.(1)-(4)

3.1 Heterogeneous Graph Context Representation

Formally, the embedding of each word w_e is concatenated with the embedding of its entity type w_t and the embedding of its coreference w_c as the representation of word $\mathbf{b}=[w_e:w_t:w_c]$. These sequences of word representations are in turn fed into a bidirectional long short-term memory (BiLSTM) to vectorize the input

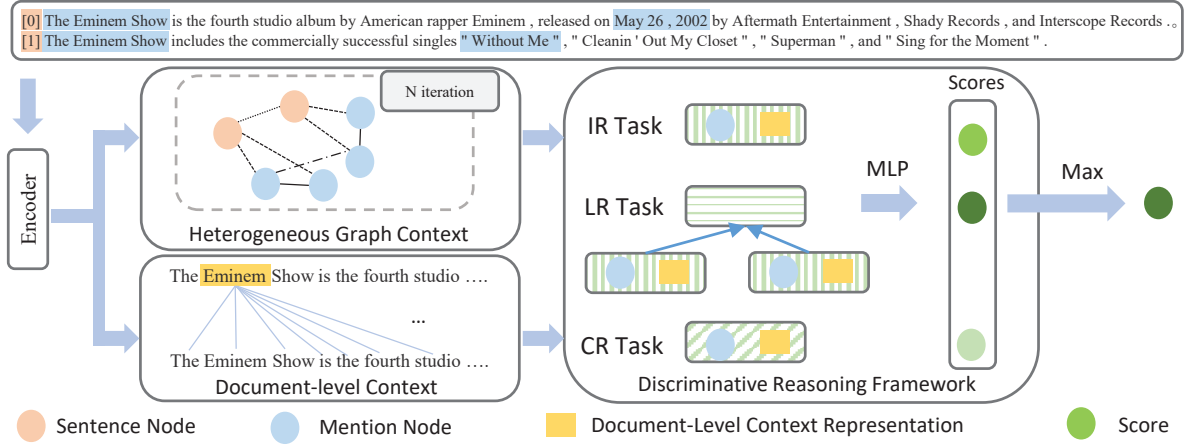


Figure 2: The overall architecture of DRN. First, A context encoder consumes the input document to get a contextualized representation of each word. Then the heterogeneous graph context representation and the document-level context representation are prepared as the input of the discriminative reasoning framework. Intra-sentence reasoning (IR) task, logical reasoning (LR) task and co-reference reasoning (CR) task are modeled explicitly and calculate the classification score respectively. Finally, the maximal score is selected as the output.

document $D = \{\mathbf{H}^1, \mathbf{H}^2, \dots, \mathbf{H}^N\}$, where $\mathbf{H}^n = (\mathbf{h}_1^n, \mathbf{h}_2^n, \dots, \mathbf{h}_{j_n}^n)$ and \mathbf{h}_i^j denotes the hidden representation of the i -th words of the j -th sentence in the document. Similar to Zeng et al.'s work, we construct a heterogeneous graph which contains sentence node and mention node. There are four kinds of edges in the heterogeneous graph: sentence-sentence edge (all the sentence nodes are connected), sentence-mention edge (the sentence node and the mention node which resides in the sentence), mention-mention edge (all the mention nodes which are in the same sentence) and co-reference edge (all the mention nodes which refer to the same entity). Then we apply the graph-based DocRE method (Zeng et al., 2020) to encode the heterogeneous graph, based on which the heterogeneous graph context representation (**HGCRep**) are learned. The **HGCRep** of each mention node and sentence node \mathbf{g}_n is formally denoted as:

$$\mathbf{g}_n = [\mathbf{v}_n : \mathbf{p}_n^1 : \mathbf{p}_n^2 : \dots : \mathbf{p}_n^{l-1}], \quad (7)$$

where $\mathbf{g}_n \in \mathbb{R}^{d_1}$ and “:” is the concatenation of vectors and each of $\{\mathbf{p}_n^1, \mathbf{p}_n^2, \dots, \mathbf{p}_n^{l-1}\}$ is learned by the multi-hop graph convolutional network (Zeng et al., 2020) and \mathbf{v}_n is the initial representation of the n -th node extracted from \mathbf{D} . Finally, there is a heterogeneous graph representation $\mathbf{G} = \{\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_N\}$ including each mention nodes and sentence nodes.

3.2 Document-level Context Representation

In the DocRE task, these reasoning skills heavily rely on the original document context information rather than the heterogeneous graph context information. Thus, the existing advanced DocRE models use syntactic trees or heuristics rules to extract the context information (i.e., entities, mentions, and sentences) that is directly related to the relation between entity pairs. However, this approach destroys the original document structure, which is weak in modeling the reasoning between two entities for the DocRE task. Therefore, we use the self-attention mechanism (Vaswani et al., 2017) to learn a document-level context representation (**DLCRep**) \mathbf{c}_n for one mention based on the vectorized input document \mathbf{D} :

$$\mathbf{c}_n = \text{softmax}\left(\frac{\mathbf{h}_j^n \mathbf{K}^\top}{\sqrt{d_{model}}}\right) \mathbf{V}, \quad (8)$$

where $\mathbf{c}_n \in \mathbb{R}^{d_2}$ and $\{\mathbf{K}, \mathbf{V}\}$ are key and value matrices that are transformed from the vectorized input document \mathbf{D} using a linear layer. Here, inspired by relation learning (Baldini Soares et al., 2019), we use the hidden state of the head word in one mention or one sentence to denote them for simplicity.

3.3 Modeling of Reasoning Paths

In this section, we use the concatenation operation to model the reasoning step on the reasoning path, thereby modeling the defined reasoning paths

in Section 2.1 as the corresponding reasoning representations as follows:

- 1) For the intra-sentence reasoning path, both HGCREps and DLCREps of two mentions are concatenated in turn as a reasoning representation:

$$\alpha_{ij} = [\mathbf{g}_{m_i^{s_1}} : \mathbf{g}_{m_j^{s_1}} : \mathbf{c}_{m_i^{s_1}} : \mathbf{c}_{m_j^{s_1}}], \quad (9)$$

where $\alpha_{ij} \in \mathbb{R}^{2d_1+2d_2}$ and “:” is the concatenation of vectors.

- 2) For the logical reasoning path, both HGCREps of mention $m_i^{s_1}$ and $m_j^{s_2}$ and DLCREps of two mention pair $(m_i^{s_1}, m_l^{s_1})$ and $(m_j^{s_2}, m_l^{s_2})$ are concatenated as their reasoning representation:

$$\beta_{ij} = [\mathbf{g}_{m_i^{s_1}} : \mathbf{g}_{m_j^{s_2}} : \mathbf{c}_{m_i^{s_1}} + \mathbf{c}_{m_l^{s_1}} : \mathbf{c}_{m_j^{s_2}} + \mathbf{c}_{m_l^{s_2}}], \quad (10)$$

where $\beta_{ij} \in \mathbb{R}^{2d_1+2d_2}$.

- 3) For the coreference reasoning path, we connect both HGCREps of two mentions and DLCREps of two sentences are concatenated in turn as their reasoning representation:

$$\gamma_{ij} = [\mathbf{g}_{m_i^{s_1}} : \mathbf{g}_{m_j^{s_2}} : \mathbf{c}_{s_1} : \mathbf{c}_{s_2}] \quad (11)$$

where $\gamma_{ij} \in \mathbb{R}^{2d_1+2d_2}$ and both \mathbf{c}_{s_1} and \mathbf{c}_{s_2} denote DLCREps for two sentences s_1 and s_2 .

The learned reasoning representations α_{ij} , β_{ij} , and γ_{ij} is as the input to classifier to compute the probabilities of relation between e_i and e_j entities by a multilayer perceptron (MLP) respectively:

$$P(r|e_i, e_j, D) = \max[\text{sigmoid}(\text{MLP}_r(\alpha_{ij}), \text{sigmoid}(\text{MLP}_r(\beta_{ij}), \text{sigmoid}(\text{MLP}_r(\gamma_{ij}))]. \quad (12)$$

Similarly, when there are multiple reasoning paths between two entities for one reasoning type in Eq.6, Eq.12 is rewritten as follows:

$$P(r|e_i, e_j, D) = \max[\text{MLP}_r(\alpha_{ij}^1), \dots, \text{MLP}_r(\alpha_{ij}^K), \text{MLP}_r(\beta_{ij}^1), \dots, \text{MLP}_r(\beta_{ij}^K), \text{MLP}_r(\gamma_{ij}^1), \dots, \text{MLP}_r(\gamma_{ij}^K)]. \quad (13)$$

Also, the binary cross-entropy is used as training objection, which is the same as the advanced DocRE model (Yao et al., 2019).

4 Experiments

4.1 Data set and Setup

Hyperparameter	Value
Batch Size	12
Optimizer	AdamW
Learning Rate	1e-3
Activation Function	ReLU
Word Embedding Size	100
Entity Type Embedding Size	20
Coreference Embedding Size	20
Encoder Hidden Size	128
Dropout	0.5
Layers of GCN	2
Weight Decay	0.0001
Device	GTX 1080Ti

Table 1: Settings for DRN.

The proposed methods were evaluated on a large-scale human-annotated dataset for document-level relation extraction (Yao et al., 2019). DocRED contains 3,053 documents for the training set, 1,000 documents for the development set, and 1,000 documents for the test set, totally with 132,375 entities, 56,354 relational facts, and 96 relation types. More than 40% of the relational facts require reading and reasoning over multiple sentences. For more detailed statistics about DocRED, we recommend readers to refer to the original paper (Yao et al., 2019).

Following settings of Yao et al.’s work, we used the GloVe embedding (100d) and BiLSTM (128d) as word embedding and encoder. The number of the reasoning path for each task is set to 3. The learning rate was set to 1e-3 and we trained the model using AdamW (Loshchilov and Hutter, 2019) as the optimizer with weight decay 0.0001 under Pytorch (Paszke et al., 2017). For the BERT representations, we used uncased BERT-Based model (768d) as the encoder and the learning rate was set to $1e^{-5}$. For evaluation, we used F1 and Ign F1 as the evaluation metrics. Ign F1 denotes F1 score excluding relational facts shared by the training and development/test sets. In particular, the predicted results were ranked by their confidence and traverse this list from top to bottom by F1 score on development set, and the score value corresponding to the maximum F1 is picked as threshold θ . The hyper-parameter for the number of reasoning paths was tuned based on the

development set. In addition, the results on the test set were evaluated through CodaLab¹. Once a model is trained, we get the confidence scores for every triple example (subject,object,relation) as Eq.(12). We rank the predicted results by their confidence and traverse this list from top to bottom by F1 score on development set, the score value corresponding to the maximum F1 is picked as threshold θ . This threshold is used to control the number of extracted relational facts on the test set.

4.2 Baseline Systems

We reported the results of the recent graph-based DocRE methods as the comparison systems: GAT (Veličković et al., 2018), GCNN (Sahu et al., 2019), EoG (Christopoulou et al., 2019), AG-GCN (Guo et al., 2019), LSR (Nan et al., 2020), GAIN (Zeng et al., 2020), and HeterGASN-Rec(Xu et al., 2021). Moreover, pre-trained models like BERT (Devlin et al., 2019) has been shown impressive result on the DocRE task. Therefore, we also reported state-of-the-art graph-based DocRE models with pre-trained BERT_{base} model, including Two-Phase+BERT_{base} (Wang et al., 2019), LSR+BERT_{base} (Nan et al., 2020), GAIN+BERT_{base} (Zeng et al., 2020), HeterGASN-Rec+BERT_{base} (Xu et al., 2021), and ATLOP-BERT_{base} (Zhou et al., 2021).

4.3 Main Results

Table 2 presents the detailed results on the development set and the test set for the DocRE dataset. First, the proposed DRN model significantly outperformed the existing graph-based DocRE systems. Second, the proposed DRN model was superior to all the existing graph-based DocRE systems on the test set, validating that modeling reasoning discriminatively is more beneficial to DocRE than the original universal neural network way. Meanwhile, it also outperformed the best HeterGASN-Rec model by 1.10 points in terms of F1, validating the effectiveness of our discriminative reasoning method. Third, for the comparisons with a pre-trained language model (BERT_{base}), F1 scores of the proposed DRN+BERT_{base} model was higher than that of the existing graph-based DocRE ATLOP+BERT model systems with BERT_{base} on the test set. In particular, our method (F1 61.37) was superior to

¹<https://competitions.codalab.org/competitions/20717>

the existing best ATLOP+BERT model (F1 61.30) in terms of F1, which is a new state-of-the-art result on the DocRE dataset.

Methods	Dev		Test	
	Ign F1	F1	Ign F1	F1
<i>Existing DocRE Systems</i>				
GCNN [†]	46.22	51.52	49.59	51.62
EoG [†]	45.94	52.15	49.48	51.82
GAT [†]	45.17	51.44	47.36	49.51
AGGCN [†]	46.29	52.47	48.89	51.45
LSR*	48.82	55.17	52.15	54.18
GAIN*	53.05	55.29	52.66	55.08
HeterGASN-Rec*	54.27	56.22	53.27	55.23
BERT _{base} *	-	54.16	-	53.20
Two-Phase BERT _{base} *	-	54.42	-	53.92
LSR+BERT _{base} *	52.43	59.00	56.97	59.05
GAIN+BERT _{base} *	59.14	61.22	59.00	61.24
HeterGASN-Rec+BERT _{base} *	58.13	60.18	57.12	59.45
ATLOP-BERT _{base} *	59.22	61.09	59.31	61.30
<i>Our DocRE Systems</i>				
DRN	54.61	56.49	54.35	56.33
DRN+BERT_{base}	59.33	61.39	59.15	61.37

Table 2: Results on the development set and the test set. Results with * are reported in their original papers. Results with † are reported in (Nan et al., 2020). Bold results indicate the best performance of the current method.

4.4 Evaluating Hyper-parameter K for The Number of Reasoning Paths

K	Dev Set		Test Set		Cover (%)
	Ign F1	F1	Ign F1	F1	
1	54.04	55.94	53.83	55.81	63.05
2	54.63	56.47	54.12	56.07	82.17
3	54.61	56.49	54.35	56.33	90.40
4	54.52	56.34	54.06	55.93	95.22
>4	54.31	56.25	53.97	55.84	100

Table 3: The effect of the number of reasoning paths K for the proposed DRN model.

To evaluate the effect of the number of reasoning path K in Eq.6, we reported the results for the different number of reasoning path K , as shown in Table 3. When K increased from 1 to 3, F1 scores of the proposed DRN model gradually improved from 55.81 to 56.33 on the test set and the percentage of covered reasoning paths reaches 90.40%. As the hyper-parameter K continues to increase, F1 scores began to drop on the dev and test sets. On the one hand, the reason may be that the reasoning information provided by too many reasoning paths is duplicated, even noises in the remaining 9.60% reasoning paths. On the other hand, the hyper-parameter $K=3$ can

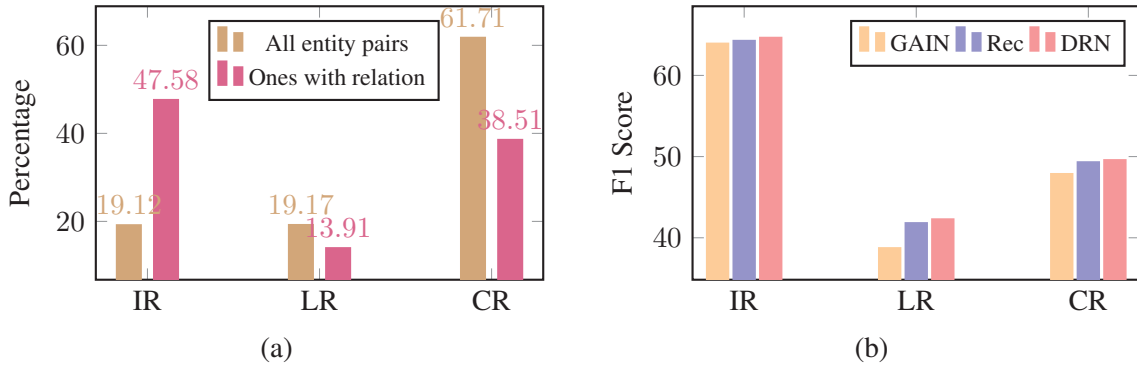


Figure 3: (a) The statistical result of different reasoning task. (b) The performance of different reasoning task.

make the proposed DRN gain the highest F1 score on the dev and test sets. Therefore, we set the hyper-parameter K to three in our main results in Table 2.

4.5 Ablation Experiments

In the proposed DRN model, we model different reasoning tasks discriminatively using HGCRep and DLCRep, and we choose the highest scores as the final results. Instead of using the discriminative reasoning framework, previous work averaged the mention representation (HGCRep or DLCRep) to get the entity representation and concatenate the two entity representation to classify the relation, which we denote as Uniform model. Table 4 shows ablation experiments of the framework and different reasoning context on the test set. It is noted that Uniform model with the discriminative reasoning framework is our DRN model. First, the DocRE models benefit from our discriminative reasoning framework no matter what the reasoning context is used. Specially, the F1 score of the model with the framework was averagely 1.21 points superior to the Uniform model on the test set no matter what context representation is used, which illustrated the effectiveness of the framework. Second, when we gradually remove

Model	without framework		with framework		Delta
	Ign F1	F1	Ign F1	F1	
Uniform	53.68	55.79	54.35	56.33	+0.83
-DLCReps	51.82	53.83	52.96	55.01	+1.33
-HGCReps	51.21	53.36	52.35	54.13	+0.97
-Both	44.73	51.06	50.68	52.78	+1.71

Table 4: Ablation experiments.

DLCRep and HGCRep from the Uniform and

the proposed DRN model, both of the model’s performance drops. Specially, F1 scores of DRN without DLCRep dropped by 1.32 while F1 scores of DRN without HGCRep dropped by 2.20 respectively. This indicates that both DLCRep and HGCRep play an important role in capturing the information of nodes on the reasoning paths. When removing both of DLCReps and HGCReps from the DRN model, the model was degraded to the BiLSTM model with our discriminative reasoning framework. Obviously, F1 scores drastically decreased on the test sets, confirming the necessity of learning DLCRep and HGCRep for modeling reasoning discriminatively.

4.6 Analysis of the Reasoning Tasks

In this section, we first showed the percent of all entity pairs (396,790) and entity pair with relation (12,332) on the dev set selected for three defined reasoning tasks through max operation in Eq.(12), as shown in Figure 3(a). For example, IR, LR, and CR are the intra-Sentence reasoning task, the logical reasoning task, and the coreference reasoning task, respectively. The percentages of IR, LR, and CR which is selected for all the entity pair are 19.12%, 19.17%, and 61.71% for all entity pairs, respectively. This indicates that our defined three reasoning skills can completely cover all entity pairs regardless of whether these entity pairs have relationships or not. Also, the percentages of IR, LR, and CR are 47.58%, 13.91%, and 38.51% for entity pairs with relation, respectively. This is consistent with the statistical result in the Yao et al.’s work that more than 40.7% relational facts can only be extracted from multiple sentences, validating that our method can model different reasoning skills discriminatively on the DocRE dataset.

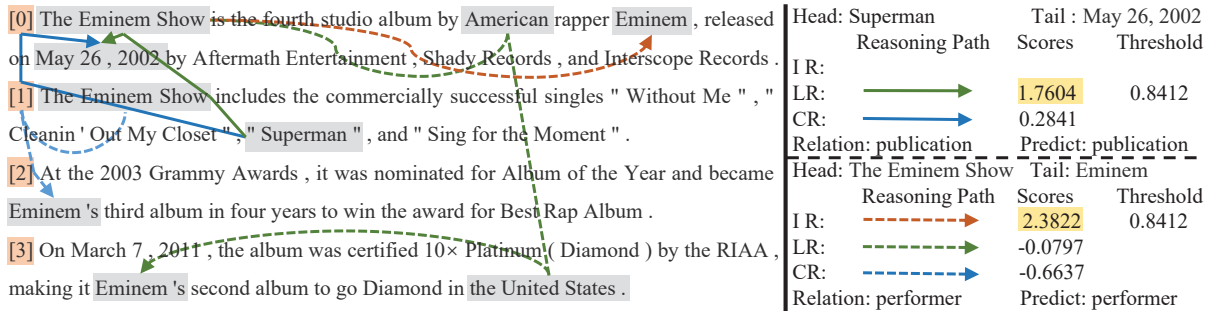


Figure 4: Case study.

Moreover, Figure 3(b) showed the results of HerterGSAN-Rec (abbreviated as Rec), GAIN, and our DRN models on three different reasoning tasks. As seen, F1 scores of the proposed DRN model are higher than that of Rec and GAIN models over all three tasks. This means that modeling reasoning types explicitly can effectively advance the DocRE. For all DocRE models, F1 scores of LR task and CR task were far inferior to that of IR task, which is consistent with the intuitive perception that the inter-sentence reasoning is more difficult than the intra-sentence reasoning.

4.7 Analysis of the Reasoning Type

		Confusion Matrix			Total
		Truth			
Predict	IR	321	36	12	369
	LR	25	88	29	142
	CR	88	129	188	405
	Total	434	253	229	
		Metric			
		IR	LR	CR	
F1 Scores		79.95	38.77	44.87	

Table 5: Confusion matrix of different reasoning types.

To further show the selected different reasoning types in Eq.(12), we randomly sampled 72 documents from the dev set which contain 916 relation instances, and we ask three human to annotate the reasoning types of all the entity pairs with relation in the sampled document according to three defined reasoning types, including the intra-sentence reasoning, the logical reasoning, and the coreference reasoning (The annotation data can be found in <https://github.com/xwjim/DRN>). Table 5 shows the number and F1 scores of each selected reasoning types on the sampled 72 documents. As seen, F1 scores of IR,

LR, and CR are 79.95%, 38.77%, and 44.87%, respectively, indicating that modeling reasoning discriminatively is working during selecting of reasoning paths in Eq.(12). Also, our method is the capacity of recognizing not only the intra-sentence reasoning but also the intra-sentence reasoning. In addition, there is a certain percentage of the mistakenly selected reasoning types, indicating that our method may have more room for improvement in the future.

4.8 Case Study

Figure 4 shows the relation classification about two entity pairs for our DRN model. For the first entity pair {"Superman"} and {"May 26,2002"}, there are reasoning paths for Task2 and Task3, and their scores are 1.7604, and 0.2841, respectively. As a result, Task2 was used to predict the relation "{publication date}" between {"Superman"} and {"May 26,2002"} correctly. Meanwhile, the selection of Task2 is consistent with the ground-truth logical reasoning type. Moreover, the above reasoning processing is also similar to the entity pair {"The Eminem show"} and {"Eminem"} with three reasoning types.

5 Related Work

Early research efforts on relation extraction concentrate on predicting the relation between two entities with a sentence (Zeng et al., 2014, 2015; Wang et al., 2016; Sorokin and Gurevych, 2017; Feng et al., 2018; Song et al., 2019; Wei et al., 2020). These approaches do not consider interactions across mentions and ignore relations expressed across sentence boundaries. The semantics of a document context is coherent and a part of relation can only be extracted among sentences.

However, as large amounts of relationships are expressed by multiple sentences, recent

work starts to explore document-level relation extraction. People begin to consider the relation between disease and chemicals in the entire document of biomedical domain (Quirk and Poon, 2017; Gupta et al., 2019; Zhang et al., 2018; Christopoulou et al., 2019; Zhu et al., 2019). A large-scale general-purpose dataset for DocRE constructed from Wikipedia articles has been proposed in (Yao et al., 2019), which has advanced the DocRE a lot. Most approaches on DocRE are based on document graphs, which were introduced by Quirk and Poon. Specifically, they use words as nodes and construct a homogenous graph using syntax parsing tools and a graph neural network is used to capture the document information. This document graph provides a unified way of extracting the features for entity pairs. Later work extends the idea by improving neural architectures (Peng et al., 2017; Verga et al., 2018; Gupta et al., 2019) or adding more types of edges (Christopoulou et al., 2019). In the Christopoulou et al.’s work, the author construct the graph which contains different granularities (sentence, mention, entity) through co-occurrence and heuristic rule to model the graph without external tools. More recent most of the approach (Christopoulou et al., 2019; Zeng et al., 2020; Xu et al., 2021) constructs heterogeneous graph through co-occurrence and heuristic rule to model the graph without external tools. In the (Zeng et al., 2020) constructed double graphs in different granularity to capture document-aware features and the interaction between entities. In the (Xu et al., 2021) introduced a reconstructor to reconstruct the path in the graph to guide the model to learning a good node representation. Other attempts focus on the multi-entity and multi-label problems (Zhou et al., 2021). Zhou et al. proposed two techniques to solve the problems, adaptive thresholding and localized context pooling.

6 Conclusion

In this paper, we propose a novel discriminative reasoning framework to consider different reasoning types explicitly. We use meta-path strategy to extract the reasoning path for different reasoning types. Based on the framework, we propose a Discriminative Reasoning Network (DRN), in which we use both the heterogeneous graph context and the document-level context to represent

different reasoning paths. The ablation study validates the effectiveness of our discriminative framework and different modules on the large-scale human-annotated DocRE dataset. In particular, our method archives a new state-of-the-art performance on the DocRE dataset. In the future, we will explore more diverse structure information (Chen et al., 2018; Chen et al., 2020; Cohen et al., 2020) from the input document for the discriminative reasoning framework, and apply the proposed approach to other NLP tasks (Zhang et al., 2020a; Chen et al., 2020; Zhang et al., 2020b).

Acknowledgments

We are grateful to the anonymous reviewers, area chair and Program Committee for their insightful comments and suggestions. The corresponding authors are Kehai Chen and Tiejun Zhao. The work of this paper is funded by the project of National Key Research and Development Program of China (No. 2020AAA0108000).

References

- Livio Baldini Soares, Nicholas FitzGerald, Jeffrey Ling, and Tom Kwiatkowski. 2019. *Matching the blanks: Distributional similarity for relation learning*. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2895–2905, Florence, Italy. Association for Computational Linguistics.
- K. Chen, R. Wang, M. Utiyama, E. Sumita, T. Zhao, M. Yang, and H. Zhao. 2020. *Towards more diverse input representation for neural machine translation*. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28:1586–1597.
- Kehai Chen, Rui Wang, Masao Utiyama, Eiichiro Sumita, and Tiejun Zhao. 2018. *Syntax-directed attention for neural machine translation*. In *AAAI Conference on Artificial Intelligence*, pages 4792–4798, New Orleans, Louisiana, USA.
- Xiao Chen, Changlong Sun, Jingjing Wang, Shoushan Li, Luo Si, Min Zhang, and Guodong Zhou. 2020. *Aspect sentiment classification with document-level sentiment preference modeling*. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3667–3677, Online. Association for Computational Linguistics.
- Fenia Christopoulou, Makoto Miwa, and Sophia Ananiadou. 2019. *Connecting the dots: Document-level neural relation extraction with edge-oriented graphs*. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*

- and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 4925–4936, Hong Kong, China. Association for Computational Linguistics.
- William W. Cohen, Haitian Sun, R. Alex Hofer, and Matthew Siegler. 2020. Scalable neural methods for reasoning with a symbolic knowledge base. In *International Conference on Learning Representations*.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. 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, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Jun Feng, Minlie Huang, Li Zhao, Yang Yang, and Xiaoyan Zhu. 2018. Reinforcement learning for relation classification from noisy data. In *Proceedings of the AAAI Conference on Artificial Intelligence*.
- Zhijiang Guo, Yan Zhang, and Wei Lu. 2019. Attention guided graph convolutional networks for relation extraction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 241–251, Florence, Italy. Association for Computational Linguistics.
- Pankaj Gupta, Subburam Rajaram, Hinrich Schütze, and Thomas A. Runkler. 2019. Neural relation extraction within and across sentence boundaries. In *The 33th AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, pages 6513–6520. AAAI Press.
- Ilya Loshchilov and Frank Hutter. 2019. Decoupled weight decay regularization. In *International Conference on Learning Representations*.
- Guoshun Nan, Zhijiang Guo, Ivan Sekulic, and Wei Lu. 2020. Reasoning with latent structure refinement for document-level relation extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1546–1557, Online. Association for Computational Linguistics.
- Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. 2017. Automatic differentiation in pytorch. In *Conference and Workshop on Neural Information Processing Systems*.
- Nanyun Peng, Hoifung Poon, Chris Quirk, Kristina Toutanova, and Wen-tau Yih. 2017. Cross-sentence n-ary relation extraction with graph lstms. *Trans. Assoc. Comput. Linguistics*, 5:101–115.
- Chris Quirk and Hoifung Poon. 2017. Distant supervision for relation extraction beyond the sentence boundary. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 1171–1182, Valencia, Spain. Association for Computational Linguistics.
- Sunil Kumar Sahu, Fenia Christopoulou, Makoto Miwa, and Sophia Ananiadou. 2019. Inter-sentence relation extraction with document-level graph convolutional neural network. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4309–4316, Florence, Italy. Association for Computational Linguistics.
- Linfeng Song, Yue Zhang, Daniel Gildea, Mo Yu, Zhiguo Wang, and Jinsong Su. 2019. Leveraging dependency forest for neural medical relation extraction. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*.
- Daniil Sorokin and Iryna Gurevych. 2017. Context-aware representations for knowledge base relation extraction. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1784–1789, Copenhagen, Denmark. Association for Computational Linguistics.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 5998–6008. Curran Associates, Inc.
- Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph attention networks. In *International Conference on Learning Representations*.
- Patrick Verga, Emma Strubell, and Andrew McCallum. 2018. Simultaneously self-attending to all mentions for full-abstract biological relation extraction. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 872–884, New Orleans, Louisiana. Association for Computational Linguistics.
- Hong Wang, Christfried Focke, Rob Sylvester, Nilesh Mishra, and William W. J. Wang. 2019. Fine-tune bert for docred with two-step process. *ArXiv*, abs/1909.11898.

- Linlin Wang, Zhu Cao, Gerard de Melo, and Zhiyuan Liu. 2016. Relation classification via multi-level attention CNNs. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1298–1307, Berlin, Germany. Association for Computational Linguistics.
- Zhepei Wei, Jianlin Su, Yue Wang, Yuan Tian, and Yi Chang. 2020. A novel cascade binary tagging framework for relational triple extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1476–1488, Online. Association for Computational Linguistics.
- Wang Xu, Kehai Chen, and Tiejun Zhao. 2021. Document-level relation extraction with reconstruction. In *The 35th AAAI Conference on Artificial Intelligence (AAAI-21)*.
- Yuan Yao, Deming Ye, Peng Li, Xu Han, Yankai Lin, Zhenghao Liu, Zhiyuan Liu, Lixin Huang, Jie Zhou, and Maosong Sun. 2019. DocRED: A large-scale document-level relation extraction dataset. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 764–777, Florence, Italy. Association for Computational Linguistics.
- Daojian Zeng, Kang Liu, Yubo Chen, and Jun Zhao. 2015. Distant supervision for relation extraction via piecewise convolutional neural networks. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1753–1762, Lisbon, Portugal. Association for Computational Linguistics.
- Daojian Zeng, Kang Liu, Siwei Lai, Guangyou Zhou, and Jun Zhao. 2014. Relation classification via convolutional deep neural network. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 2335–2344, Dublin, Ireland. Dublin City University and Association for Computational Linguistics.
- Shuang Zeng, Runxin Xu, Baobao Chang, and Lei Li. 2020. Double graph based reasoning for document-level relation extraction. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1630–1640, Online. Association for Computational Linguistics.
- Yuhao Zhang, Peng Qi, and Christopher D. Manning. 2018. Graph convolution over pruned dependency trees improves relation extraction. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2205–2215, Brussels, Belgium. Association for Computational Linguistics.
- Zhisong Zhang, Xiang Kong, Zhengzhong Liu, Xuezhe Ma, and Eduard Hovy. 2020a. A two-step approach for implicit event argument detection. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7479–7485, Online. Association for Computational Linguistics.
- Zhuosheng Zhang, Yuwei Wu, Junru Zhou, Sufeng Duan, Hai Zhao, and Rui Wang. 2020b. Sg-net: Syntax-guided machine reading comprehension. In *Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-2020)*, pages 9636–9643.
- Wenxuan Zhou, Kevin Huang, Tengyu Ma, and Jing Huang. 2021. Document-level relation extraction with adaptive thresholding and localized context pooling. In *Proceedings of the AAAI Conference on Artificial Intelligence*.
- Hao Zhu, Yankai Lin, Zhiyuan Liu, Jie Fu, Tat-Seng Chua, and Maosong Sun. 2019. Graph neural networks with generated parameters for relation extraction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1331–1339, Florence, Italy. Association for Computational Linguistics.