Global Inference with Explicit Syntactic and Discourse Structures for Dialogue-Level Relation Extraction

Hao Fei, **Jingye Li**, **Shengqiong Wu**, **Chenliang Li**, **Donghong Ji** and **Fei Li**[*]

Key Laboratory of Aerospace Information Security and Trusted Computing, Ministry of Education, School of Cyber Science and Engineering, Wuhan University, China

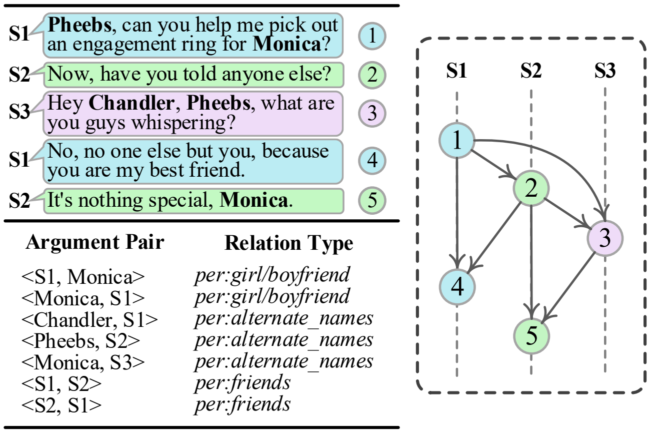
{hao.fei, theodorelee, whuwsq, cllee, dhji, lifei_csnlp}@whu.edu.cn

Abstract

Recent research attention for relation extraction has been paid to the dialogue scenario, i.e., dialogue-level relation extraction (DiaRE). Existing DiaRE methods either simply concatenate the utterances in a dialogue into a long piece of text, or employ naive words, sentences or entities to build dialogue graphs, while the structural characteristics in dialogues have not been fully utilized. In this work, we investigate a novel dialogue-level mixed dependency graph (D^2G) and an argument reasoning graph (ARG) for DiaRE with a global relation reasoning mechanism. First, we model the entire dialogue into a unified and coherent D^2G by explicitly integrating both syntactic and discourse structures, which enables richer semantic and feature learning for relation extraction. Second, we stack an ARG graph on top of D^2G to further focus on argument inter-dependency learning and argument representation refinement, for sufficient argument relation inference. In our global reasoning framework, D^2G and ARG work collaboratively, iteratively performing lexical, syntactic and semantic information exchange and representation learning over the entire dialogue context. On two DiaRE benchmarks, our framework shows considerable improvements over the current best-performing baselines. Further analyses show that the model effectively solves the long-range dependence issue, and meanwhile gives explainable predictions.

1 Introduction

Dialogue-level relation extraction is a newly proposed task that seeks to infer the semantic relationships between the subject arguments and object arguments in a conversation [Yu *et al.*, 2020], as exemplified in Fig. 1. Comparing with sentence-level RE [Katiyar and Cardie, 2016; Fei *et al.*, 2020a] and document-level RE [Yao *et al.*, 2019], DiaRE is much more challenging due to the characteristic of dialogues. Existing DiaRE studies handle multi-turn dialogues by concatenating all the utterances within it as a very long text sequence [Yu *et al.*, 2020; Zhao *et al.*, 2021], while recent studies model the dialogue context as various graphs so as to learn better feature representations [Chen *et al.*, 2020; Xue *et al.*, 2021; Nan *et al.*, 2021; Qiu *et al.*, 2021]. Unfortunately, current research still fails to dig into several pivotal issues of DiaRE, which hinders the task for further improvements.



S1: **Pheebs**, can you help me pick out an engagement ring for **Monica**? 1

S2: Now, have you told anyone else? 2

S3: Hey **Chandler**, **Pheebs**, what are you guys whispering? 3

S1: No, no one else but you, because you are my best friend. 4

S2: It's nothing special, **Monica**. 5

Argument Pair	Relation Type
<S1, Monica>	*per:girl/boyfriend*
<Monica, S1>	*per:girl/boyfriend*
<Chandler, S1>	*per:alternate_names*
<Pheebs, S2>	*per:alternate_names*
<Monica, S3>	*per:alternate_names*
<S1, S2>	*per:friends*
<S2, S1>	*per:friends*

Figure 1: **Left**: dialogue-level relation extraction. **Right**: dialogue-answering structure.

First of all, multi-party dialogue threads are scattered and entangled, and the semantics or topic consistency of each thread will be interrupted and damaged when simply concatenating the utterances into a long flat sequence. In fact, there could be a logical answering structure between utterances from different speakers (parties), as illustrated in the right part of Fig. 1. **Second**, the key to infer the relation of argument pairs lies in locating the crucial trigger clues in texts, for which the syntactic dependency tree features have been extensively and successfully exploited in regular RE [Miwa and Bansal, 2016; Fei *et al.*, 2020d; Fei *et al.*, 2020c]. However, the overall dialogue contexts are segmented into utterance pieces, which makes it intractable to directly apply the syntax structure information for DiaRE. **Third**, the speaker coreference ambiguity is not fully investigated in existing studies, i.e., the issue of *first-person* ('I'), *second-person* ('you') and *zero pronoun* of speakers in utterances would hinder the relation inference. **Fourth**, there

[*] Corresponding author

could be complex relation inter-dependencies between the arguments in DiaRE, such as the *multi-hop relations*, *implicit relations*, *reversed relations*, etc. This requires an effective method for global-level argument reasoning.

In this work, we address all the above challenges for improving DiaRE. First, we model the entire conversation text into a hierarchical **dialogue-level mixed dependency graph** (D^2G). As demonstrated in Fig. 2, the D^2G is a directed acyclic graph by connecting 1) the inter-utterance structures including *dialogue answering network* & *speaker coreference links*, and 2) the intra-utterance structures including *syntactic dependency tree* & *speaker-predicate links*. D^2G organizes the entire dialogue into a coherent dependency graph that explicitly integrates both the syntactic and discourse structures, which enables to more accurately capture the dialogue semantics and critical trigger clues for relation inference.

On the other hand, we perform end-to-end DiaRE with global relation reasoning. Conditioned on the argument mentions from D^2G, we build a bidirectional **argument reasoning graph** (ARG) for direct argument relation inference. Overall, the system consistently refines argument representations at the global level based on the dual graph (i.e., both D^2G and ARG), and finally outputs all the predicted relations between argument pairs via a triaffine decoder (cf. Fig. 3). Over both the English and Chinese DiaRE datasets, our system outperforms the current state-of-the-art (SoTA) models with big margins. Further analyses show the importances of constructing dialogue-level dependency structures and the global relation reasoning mechanism for DiaRE.

To sum up, this paper contributes mainly in three folds.

★ We introduce a novel dialogue-level mixed dependency graph, D^2G, which integrates syntactic and discourse structural information from various aspects. D^2G enhances the overall semantic learning of dialogue contents and the feature retrieval of argument pairs.

★ We introduce an argument reasoning graph, ARG, for direct inference of the argument inter-dependencies. We aggregate the argument mentions in ARG from D^2G via a conditional argument node normalization mechanism.

★ Our framework achieves new SoTA performances on benchmarks, and meanwhile yields explainable predictions.[1]

2 Related Work

Relation extraction (RE) has long been a fundamental NLP task, aiming at discovering argument relations in given texts [Katiyar and Cardie, 2016]. RE was upgraded from the initial sentence level to the document level, which recently has been introduced at the dialogue scenario, i.e., DiaRE [Yu *et al.*, 2020]. The crux of RE is to deeply understand the context semantics and accurately retrieve the critical features for revealing the relations of argument pairs. Comparing with sentence-level and document-level RE, the relation inference in DiaRE could be much more difficult because of the nature of conversation form of texts, e.g., non-sequential order of discourse structure, scattered clues in different utterances, and speaker coreference.

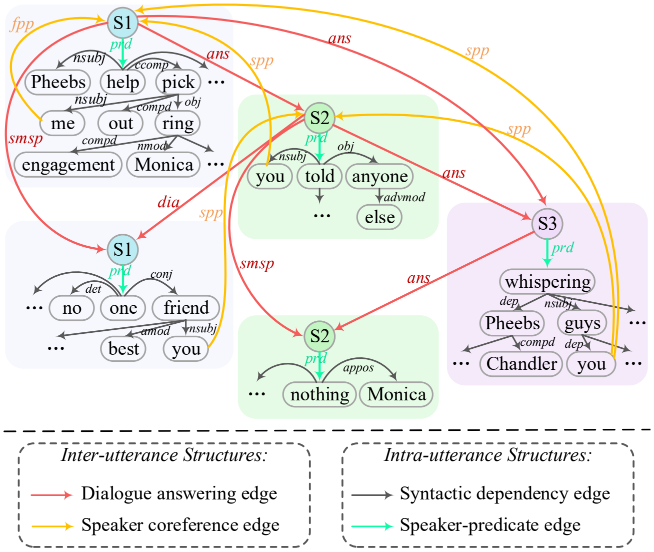


Figure 2: Dialogue-level mixed dependency graph (D^2G).

As a newly proposed task, currently DiaRE has received limited research attention. Initial DiaRE works [Yu *et al.*, 2020; Zhao *et al.*, 2021; Long *et al.*, 2021] directly collapse the multi-turn dialogues into a document, i.e., transforming into the document-level RE. This however would disrespect the aforementioend conversation characteristics, and thus lead to suboptimal results. Very recent works consider constructing graph representations for DiaRE. For example, Chen et al., [2020] build a graph that connects the speaker, entity, type, and utterance nodes. Xue et al., [2021] and Nan et al., [2021] develop the latent graphs for DiaRE so as to better capture the key features for relation inference. Although improvements have been made, existing graph-based methods still fail to fully leverage the dialogue structural features, i.e., the dialogue discourse structure and the inner-utterance structure as we revealed previously.

This work also closely relates to the syntactic dependency-based RE methods [Xu *et al.*, 2015; Fei *et al.*, 2020b]. The external dependency structure provides intrinsic prior knowledge for mining the critical feature for relation inference from a low-level linguistic perspective [Fei *et al.*, 2021b; Wu *et al.*, 2021], which thus significantly promotes the RE performances [Miwa and Bansal, 2016; Song *et al.*, 2019]. However, the syntax dependency information could not be directly applied to DiaRE task, as the dialogue contexts are segmented into utterance pieces with arbitrary order. In this work, we construct a novel hierarchical dialogue-level mixed dependency graph for DiaRE, representing the overall conversation as a coherent structure.

3 Dialogue-level Mixed Dependency Graph

We model a dialogue as a unified directed acyclic structure, i.e., dialogue-level mixed dependency graph as illustrated in Fig. 2. We formulate D^2G as G=(V, E), where V is a set of nodes v_i of words and speakers, and E is a set of labeled edges $\pi_{u,v}$, with E=E_{ans}∪E_{sco}∪E_{dep}∪E_{spd}. E_{ans} repre-

[1] Codes at https://github.com/scofield7419/DiaRE-D2G.

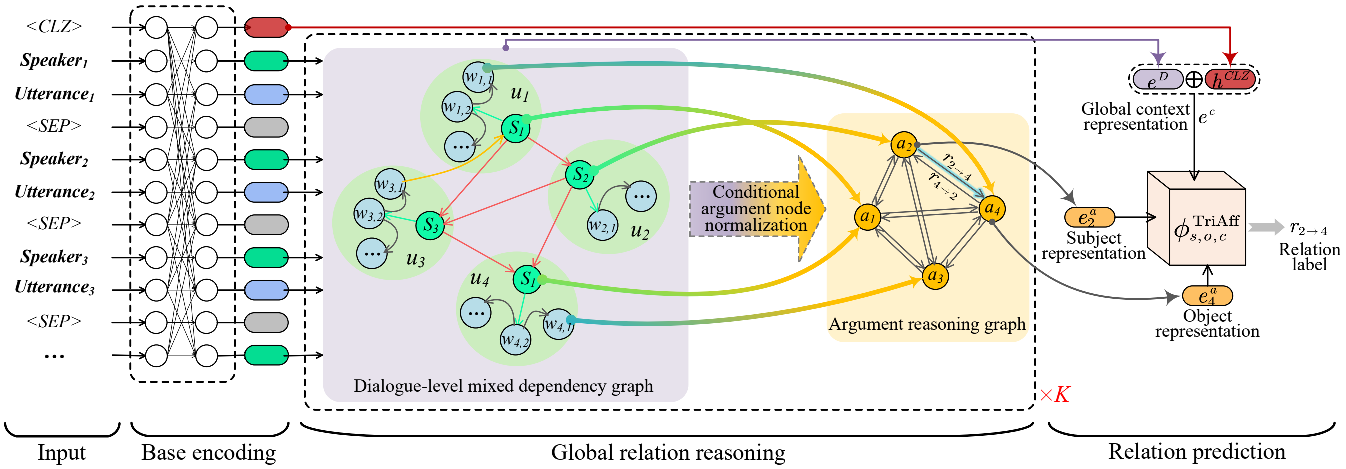


Figure 3: The overall DiaRE framework, which consists of four tiers. First, the base encoder generates contextual representations for the input dialogue texts. Then, the global relation reasoning module performs argument relation inference over the dialogue via the D^{2}G and the ARG. Finally, a triaffine decoder carries out relation prediction for all the argument pairs end-to-end.

sents the dialogue answering edges between different utterances; E_{sco} represents the speaker-coreference edges; E_{dep} represents the sentence-level syntactic dependency edges; E_{spd} represents the speaker-predicate edges. The former two types of edges refer to the inter-utterance structures while the latter two types of edges refer to the intra-utterance structures.

Dialogue answering edge. The dialogue answering structure can be seen as an inter-utterance dialogue discourse structure, ensuring a information flow from predecessors to successor with semantics consistency. Specifically, each conversational utterance u_i will connect to a subsequent utterance u_j ($i<j$) that either u_j is a response to u_i (i.e., cross-speaker case), or both u_i and u_i are yielded from a same speaker. For the cross-speaker case which is essentially a multi-turn response selection problem [Lu *et al.*, 2019; Jia *et al.*, 2020], we employ a well-trained off-the-shelf model to predict to which previous utterance the current one should link, and assign an '*ans*' (answer) label for the edge. Also we directly link these utterances with same speakers and with a '*smsp*' (same speaker) label.

Speaker coreference edge. Speaker coreference information should also be properly modeled. First, for the case of first-person pronoun of speaker (e.g., 'I', 'me', 'myself'), we directly link those pronoun words to the utterance speaker node with edge type '*fpp*' (first-person pronoun). Then, for the second-person speaker pronoun (e.g., 'you'), we create '*spp*' type of edges from the pronoun to the corresponding speaker(s) in the adjacent utterance(s) determined by the aforementioned '*ans*' arcs. Note that here we do not consider the third-person pronoun case, due to its particular difficulty for disambiguation.

Syntactic dependency edge. We represent the inner-utterance sentences by the syntactic dependency trees that are produced from a external third-party parser. In the syntactic dependency tree, each dependency edge links the head word to the dependent word with a specific syntactic label, e.g., 'Pheebs'$\xleftarrow{nsubj}$'help' shown in Fig. 2.

Speaker-predicate edge. We then connect the speaker to its corresponding utterance, so as to make the speaker&utterance a coherent unit, and meanwhile solve the zero pronoun issue (omitted reflexive pronoun). Technically, we create the speaker-predicate edges that an utterance speaker will link to the core predicate word of the utterance (with edge type of '*prd*').[2] When an utterance contains multiple sentences, we create multiple speaker-predicate edges, i.e., many predicates to one speaker projection.

4 DiaRE Framework

Task formalization. In DiaRE, a dialogue includes a sequence of utterances $\{u_n\}_{n=1}^{N}$, and a set of argument pairs A=$\{(a_i, a_j)_o\}_{o=1}^{|A|}$. Each utterance is a sequence of words u_n=$\{w_{n,1}, \cdots, w_{n,m}\}$, yielded by a corresponding speaker $S_l \in \{S_l\}_{l=1}^{L}$. An argument could either be an entity mention in utterances or a speaker.[3] Our system also creates a D^{2}G for the corresponding dialogue texts. The target is to predict the relation label $r_{i\rightarrow j} \in R$ between the subject argument a_i and the object argument a_j.[4] We also include a dummy label ϵ in R to represent no valid relation between a_i and a_j.

4.1 Base Encoding

Following the line of DiaRE works [Yu *et al.*, 2020; Nan *et al.*, 2021; Long *et al.*, 2021], we also employ the pre-trained BERT language model [Devlin *et al.*, 2019] as the underlying encoder to yield the contextualized representations for the words and speakers. We pack the utterance with its speaker as a group, and concatenate those groups into a whole (separated with *SEP* tokens), and feed into BERT encoder:

$$\begin{aligned} X &= \{CLZ, (s_l, w_{n,1}, \cdots, w_{n,m}), SEP, \cdots\}\,, \\ \{(\boldsymbol{h}_l^s, \boldsymbol{h}_{n,1}^w, \cdots, \boldsymbol{h}_{n,m}^w)\}_{n=1}^{N} &= \mathrm{BERT}(X)\,, \end{aligned} \tag{1}$$

[2] We note that mostly the core predicate of a sentence is the only child of the virtual *Root* node within the syntax tree.

[3] 89.9% of argument pairs have at least one speaker in the dataset.

[4] The relation is directional, i.e., $r_{i\rightarrow j} \neq r_{j\rightarrow i}$.

where $\boldsymbol{h}^s_l$ is the representation of speaker s_l, and $\boldsymbol{h}^w_{n,*}$ is word representation, both of which will be used in the next module.

4.2 Global Relation Inference

Global relation reasoning module performs two learning targets: 1) critical feature mining for argument pairs, which is fulfilled based on the D^{2}G encoder; 2) inter-dependencies inference for argument relations via the ARG encoder. Both of two graphs works collaboratively, performing the relation reasoning globally over the dialogue.

D^{2}G encoding. Each D^{2}G edge comes with a label. To encode D^{2}G effectively, we here propose a novel label-wise graph convolutional network (LWGCN). In the graph $G = (V, E)$, for each edge $\pi_{i,j} \in E$ from node v_i to v_j, we define $\pi_{i,j}$=1 when there is an edge in between, and $\pi_{i,j}$=0 vice versa. We additionally add a '*self*' label as the self-loop arc (i.e., $\pi_{i,i}$=1) for each node v_i to enrich the information aggregation. We also maintain the vectorial embedding $\boldsymbol{x}^\pi_{i,j}$ for each edge label. We denote the LWGCN hidden representation of node v_i as $\boldsymbol{e}_i$:

$$\boldsymbol{e}^d_i = \mathrm{ReLU}(\textstyle\sum_j \gamma_{i,j}(\boldsymbol{W}_1 \cdot \boldsymbol{h}_j + \boldsymbol{W}_2 \cdot \boldsymbol{x}^\pi_{i,j} + b))\,, \tag{2}$$

where $\boldsymbol{h}_j$ is the node representation from BERT encoder (cf. Eq. 1), and $\gamma_{i,j}$ is the linking distribution calculated via:

$$\boldsymbol{e}^s_{i,j} = \boldsymbol{W}_3 \cdot [\boldsymbol{h}_j; \boldsymbol{x}^\pi_{i,j}]\,, \tag{3}$$

$$\gamma_{i,j} = \frac{\pi_{i,j} \cdot \exp(\boldsymbol{e}^s_{i,j})}{\sum_z \pi_{i,z} \cdot \exp(\boldsymbol{e}^s_{i,z})}\,. \tag{4}$$

$\gamma_{i,j}$ indicates the structural neighboring connecting strength globally, which will be dynamically updated during learning so that some important clues will be highly weighted and lead to more accurate relation detection.

ARG encoding. In DiaRE, the direct information exchanging of different arguments should be considered for sufficient relation inference (e.g., argument inter-dependencies). Thus we build an ARG, in which we create fully bidirectional connections between each argument pair, and the argument mentions are aggragated from D^{2}G,[5] as depicted in Fig. 3. We introduce a novel conditional graph neural network (ConGNN) to encode ARG. Unlike the vanilla GNN that makes direct propagation among nodes, in ConGNN, the neuron's activity of argument mention node is normalized from D^{2}G, so as to reduce the covariate shift problem that causes imbalanced inference [de Vries *et al.*, 2017; Xiong *et al.*, 2020]. Technically, ConGNN passes messages for each argument a_i as:

$$\boldsymbol{e}^a_i = \mathrm{ReLU}(\bar{D}^{-\frac{1}{2}} B \bar{D}^{-\frac{1}{2}}\, \boldsymbol{W}_4 \cdot \hat{\boldsymbol{e}}^a_i)\,, \tag{5}$$

where B is the connecting weight between an argument pair with $B_{i,j} = 1$, and $\bar{D} = \sum_v B_{i,j} = 1$. $\hat{\boldsymbol{e}}^a_i$ is the conditionally normalized node representations (ConNorm):

$$\hat{\boldsymbol{e}}^a_i = \mathrm{ConNorm}(\boldsymbol{e}^a_i, \alpha, \beta | \boldsymbol{e}^d_i) = \alpha \odot (\frac{\boldsymbol{e}^a_i - \mu}{\sigma}) + \beta\,, \tag{6}$$

$$\mu = \frac{1}{M}\textstyle\sum_j^M \boldsymbol{e}^a_{i,j}\,,\ \sigma = \sqrt{\frac{1}{M}\textstyle\sum_j^M (\boldsymbol{e}^a_{i,j} - \mu)^2}\,, \tag{7}$$

where $\boldsymbol{e}^a_{i,j}$ is the j-th element of vector $\boldsymbol{e}^a_i$, μ and σ are the mean and standard deviation of the normalization. ConNorm generates α and β by aggregating the raw mention representation $\boldsymbol{e}^d_i$ in D^{2}G:

$$\alpha = \boldsymbol{W}^\alpha \boldsymbol{z} + b^\alpha\,,\ \beta = \boldsymbol{W}^\beta \boldsymbol{z} + b^\beta\,,\ \boldsymbol{z} = \frac{1}{V}\textstyle\sum_{u=1}^V \boldsymbol{e}^d_{i,u}\,, \tag{8}$$

where V is the number of the mention representation $\boldsymbol{e}^d_{i,u}$ that refers to the same argument a_i.

Global inference with two graphs. The global relation reasoning assembles the above two graph encoders as a whole, consistently performing feature learning and refining the argument relations globally. Overall, we enable total K rounds of reasoning for a sufficient information propagation.

4.3 Prediction and Training

Based on the argument representations $\boldsymbol{e}^a_i$ we finally perform end-to-end prediction for all the argument pairs. Most prior works simply concatenate two representations for prediction. However, this could inevitably lead to the order information loss between the subject and object arguments, as the DiaRE task is sensitive to the order of the argument pair. Also some global context information is not utilized in existing works. We thus employ a TriAffine decoder [Carreras, 2007] that makes decisions based on the two argument features (in order) as well as a global context feature:

$$\phi^{\mathtt{TriAff}}_{s,o,c} = \begin{bmatrix} \boldsymbol{e}^a_s \\ 1 \end{bmatrix}^{\mathrm{T}} (\boldsymbol{e}^c_o)^{\mathrm{T}} \boldsymbol{W}_5 \begin{bmatrix} \boldsymbol{e}^c \\ 1 \end{bmatrix}, \tag{9}$$

$$r_{s \to o} = \mathrm{Softmax}(\phi^{\mathtt{TriAff}}_{s,o,c})\,, \tag{10}$$

where $\boldsymbol{e}^a_s$ and $\boldsymbol{e}^o_s$ are the subject and object argument representations from ARG, $\boldsymbol{e}^c$ is global context representation:

$$\boldsymbol{e}^c = [\boldsymbol{h}^{CLZ}; \boldsymbol{e}^D]\,, \tag{11}$$

where $\boldsymbol{h}^{CLZ}$ is the BERT representation of '*CLZ*' token, $\boldsymbol{e}^D$ is the average pooling representation over all the last-layer of LWGCN node features $\{\boldsymbol{e}^d_i\}$. Only the predicted label that is valid (i.e., $r_{i \to j} \neq \epsilon$) will be output.

The training target of our system is to minimize the cross-entropy loss $\mathcal{L}$ between the predicted and ground truth labels of all the relations.

5 Experimentation

5.1 Setups

We conduct experiments on the DiaRE benchmark data [Yu *et al.*, 2020], which includes the English version (DialogRE-EN) and the Chinese translation version (DialogRE-CN). DiaRE data is split into Train&Dev&Test sets, and totally contains 1,788 dialogues and 10,168 relational triples, covering 36 relation types, with average of 13.1 utterance per dialogue and average 3.3 speaker per utterance. To yield dialogue answering edges E_{ans}, we adopt the current SoTA multi-turn response selection model [Jia *et al.*, 2020]. We employ the Stanford CoreNLP Toolkit[6] to obtain the dependency parse trees E_{dep}. We load the base version BERT parameters.

[5] One argument entity is often mentioned more than once in D^{2}G, and scattered broadly within the dialogue.

[6] https://stanfordnlp.github.io/CoreNLP/, v4.2.0 typed version.

	DialogRE-EN		DialogRE-CN	
	Dev	Test	Dev	Test
RawBERT [Yu *et al.*, 2020]	63.0†	61.2†	65.5‡	63.5‡
GDPNet [Xue *et al.*, 2021]	67.1†	64.9†	64.1‡	62.8‡
End2End [Zhou *et al.*, 2021]	65.1†	64.9†	63.9†	64.0†
AuxPrd [Zhao *et al.*, 2021]	66.8†	65.5†	-	-
HeterGraph [Chen *et al.*, 2020]	68.7†	67.4†	-	-
LatGraph [Nan *et al.*, 2021]	69.6†	68.1†	66.7†	65.4†
SocAoG [Qiu *et al.*, 2021]	69.5†	69.1†	-	-
CoIN [Long *et al.*, 2021]	71.1†	72.3†	-	-
Ours	**73.7**	**74.5**	**70.3**	**69.8**

Table 1: Results on two datasets. Values with † are copied from the corresponding raw papers; with ‡ are copied from [Nan *et al.*, 2021];

	DialogRE-EN		DialogRE-CN	
	Test	Δ	Test	Δ
Ours	**74.5**		**69.8**	
• D^{2}G				
w/o D^{2}G	67.1	-7.4	63.4	-6.4
w/o E_{ans}	68.2	-6.3	64.6	-5.2
w/o E_{sco}	72.1	-2.4	67.0	-2.8
w/o E_{dep}	67.5	-7.0	63.9	-5.9
w/o E_{spd}	73.8	-0.7	69.3	-0.5
LWGCN w/o Edge labels ($\pi_{i,j}$)	72.9	-1.6	68.5	-1.3
• ARG				
w/o ARG	71.8	-2.7	66.5	-3.3
ConGNN w/o CondNorm	72.6	-1.9	67.8	-2.0
• Prediction				
w/o Global context (e^c in Eq. 11)	72.7	-1.8	67.9	-1.9
→Concat	72.3	-2.2	67.5	-2.3

Table 2: Ablation results (F1) on two datasets.

All the BERT output representation $\boldsymbol{h}$ has 768 D. The edge label embedding ($\boldsymbol{x}^{\pi}_{i,j}$) size is 100. LWGCN hidden size, argument embedding size and ConGCN hidden size are all set as 300. We adopt the Adam optimizer with an initial learning rate of 4e-5. We set unfixed epochs with an early-stop training strategy instead. We mainly make comparisons with the existing DiaRE baselines. All the baselines use the same BERT-base embedding. We adopt the F1 score as the metric.

5.2 Results and Analyses

Main performances. In Table 1 we compare the main performances against baseline DiaRE systems. The first observation is that the RawBERT model that collapses the entire diagloue texts as a flat document also without using any other information source, presents comparatively weaker performances. In contrast, those baselines that either take the graph modeling of dialogue (i.e., HeterGraph, LatGraph), or make use of additional information (i.e., AuxPrd, End2End, Position, CoIN) achieve better results than RawBERT model.

Most importantly, our model outperforms the best-performing baselines with big margins, e.g., 2.2%(74.5-72.3) on DialogRE-EN and 3.4%(69.8-64.4) test F1 on DialogRE-CN respectively over the CoIN model. We note that CoIN is the SoTA baseline because of the design of multiple learning constraints [Long *et al.*, 2021]. However, CoIN becomes inferior to our system, largely due to the leverage of dialogue-level dependency mixed graph and the global relation reasoning mechanism in our method.

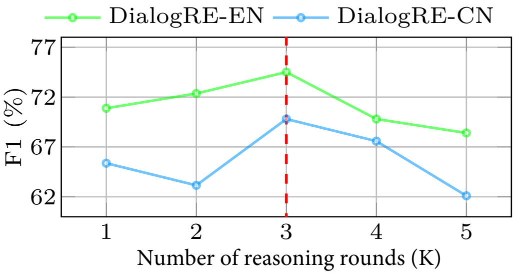


Figure 4: Influence of the round for global relation reasoning.

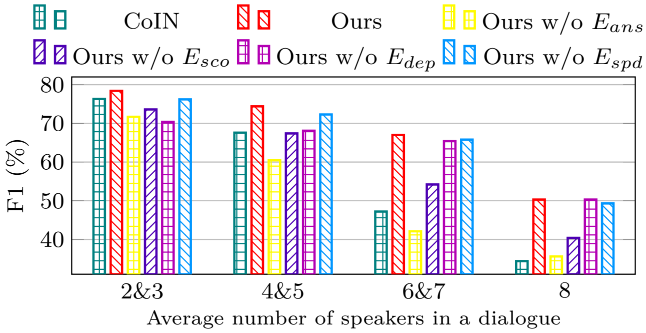


Figure 5: Influence of the speaker number in a dialogue.

Ablation study. We perform ablation experiments (cf. Table 2) to better understand the impact of each part of our proposed method. We first study the influence of the D^{2}G by removing it, which substantially results in the biggest performance drops among all the other factors, i.e., -7.4% and -6.4% F1 scores on two datasets respectively. This reflects the importances to build a dialogue-level structure for DiaRE. Diving into the D^{2}G, we remove each sub-structure one by one, and find that the syntactic dependency links are the most important element, followed by the dialogue answering edges. Besides, without encoding the edge label information, the results drop about 2 points.

Further, removing the ARG also hurts the overall performances to certain extent (not as significant as without D^{2}G). When the condition normalization mechanism of argument nodes is not available, considerable drops are witnessed, which proves the necessity of its proposal. Finally, we find that stripping off the the global context features $\boldsymbol{e}^c$ will lead to performance degradation. If further using a concatenation operation as a replacement for generating the feature representation, i.e., $[\boldsymbol{e}^a_s;\boldsymbol{e}^a_o]$ and without considering the argument order, we can meet further performance decreases.

Influence of the global reasoning round. In Fig. 4 we study how the global reasoning round affects the model performances. We see that both the performance of English and Chinese data climbs to the peak when gradually stepping into the third iteration. This informs that K=3 is enough to ensure sufficient sentiment and context learning. Once over third rounds, the overall results are deteriorated rapidly, largely due to the overfitting by too many reasoning steps.

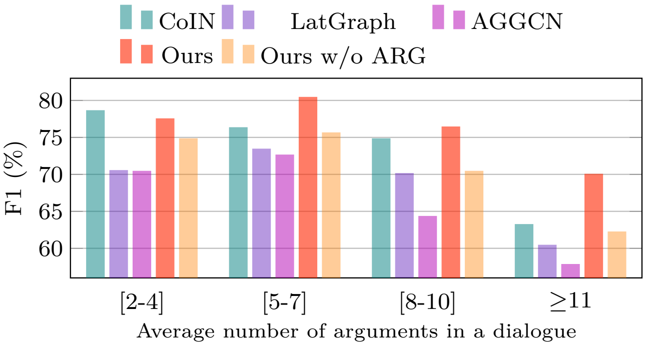


Figure 6: Influence of the argument number in a dialogue.

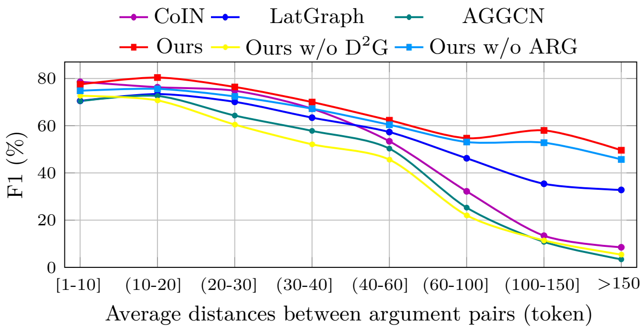


Figure 7: Influence of the distance of a argument pair.

Influence of the number of speaker parities. In multi-party dialogues, more speakers will introduce more conversation threads, i.e., more complex dialogue semantics. In Fig. 5 we show the results under different numbers of speakers. We see that without leveraging the inter-utterance structure in D^{2}G (i.e., E_{ans} and E_{sco}), our system could perform worse rapidly when handling multitudinous speakers, e.g., the speaker number is larger than 6. This evidently proves the necessity to model the cross-utterance information for DiaRE.

Influence of the relational argument number. Fig. 6 further plots the performances under different co-existed argument numbers in a dialogue. We notice that increasing the arguments causes worse overall results, since more arguments could lead to complicated relation inter-dependencies. In particular, without the integration of argument reasoning graph, the performances of our system on the bigger number of arguments (e.g., ≥11) are hurt dramatically, which demonstrates the efficacy of the ARG.

Influence of the distance of an argument pair. Sentence-level syntactic dependency tree knowledge has been extensively verified effective on relieving the long-range dependence issue in sentence-level relation extraction [Xu *et al.*, 2015; Fei *et al.*, 2020b; Wu *et al.*, 2021; Fei *et al.*, 2021a]. Here we explore the results of different systems when handling the argument pairs in different distances in DiaRE scenario. As shown in Fig. 7, we see that our system equipped with the dialogue-level dependency structure can still perform well for those super-long argument pairs, where the other baselines fail to give competitive performances (e.g., AGGCN). This implies the importance to construct the dialogue-level mixed dependency graph for DiaRE.

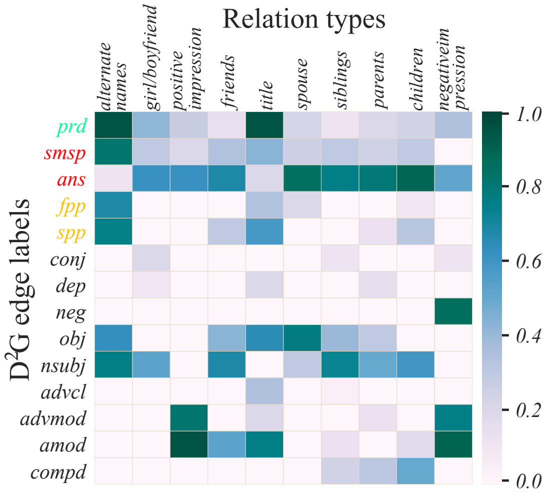


Figure 8: Discovered correlations between D^{2}G edges (left) and relation types (upper). Only a subset of the high-frequency dependency labels and argument relations is shown.

Structural correlation discovery. Finally, we qualitatively investigate if our system can capture the intrinsic correlations between the dialogue dependency structures and the argument relations. We technically observe the connecting weights $\gamma_{i,j}$ (in Eq. 4) and collect the weights of the correlated edges and the argument relation types, which are normalized and rendered in Fig. 8. Interestingly, via some patterns we can infer that our system has successfully learned some structural correlations, which accordingly explains the task improvements by our model. For example, the inter-utterance edges *dialogue answering* (‘*ans*’) show bigger influence to most of the relation types, while the relation types ‘*alternate_names*’ and ‘*title*’ rely more on the *speaker-predicate* (‘*prd*’) edges. This also reveals that our model can achieve explainable predictions for DiaRE.

6 Conclusions

In this paper, we introduce a novel system for dialogue-level relation extraction (DiaRE) task. We first propose modeling the conversation texts as a dialogue-level mixed dependency graph for more accurate feature learning, in which we integrate both syntactic and discourse information. We then introduce an argument reasoning graph with a conditional argument node normalization mechanism for direct inference of the argument inter-dependencies. Our framework achieves new state-of-the-art results over best-performing baselines on two DiaRE benchmark datasets.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No.61772378, No. 62176187), the National Key Research and Development Program of China (No. 2017YFC1200500), the Research Foundation of Ministry of Education of China (No.18JZD015), the Youth Fund for Humanities and Social Science Research of Ministry of Education of China (No. 22YJCZH064), the General Project of Natural Science Foundation of Hubei Province (No.2021CFB385). This work is also supported by the Fundamental Research Funds for the Central Universities.

References

[Carreras, 2007] Xavier Carreras. Experiments with a higher-order projective dependency parser. In *EMNLP*, pages 957–961, 2007.

[Chen *et al.*, 2020] Hui Chen, Pengfei Hong, Wei Han, Navonil Majumder, and Soujanya Poria. Dialogue relation extraction with document-level heterogeneous graph attention networks. *CoRR*, abs/2009.05092, 2020.

[de Vries *et al.*, 2017] Harm de Vries, Florian Strub, Jérémie Mary, Hugo Larochelle, Olivier Pietquin, and Aaron C. Courville. Modulating early visual processing by language. In *NeurIPS*, pages 6594–6604, 2017.

[Devlin *et al.*, 2019] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: pre-training of deep bidirectional transformers for language understanding. In *NAACL*, pages 4171–4186, 2019.

[Fei *et al.*, 2020a] Hao Fei, Yafeng Ren, and Donghong Ji. Boundaries and edges rethinking: An end-to-end neural model for overlapping entity relation extraction. *Information Processing & Management*, 57(6):102311, 2020.

[Fei *et al.*, 2020b] Hao Fei, Yafeng Ren, and Donghong Ji. Improving text understanding via deep syntax-semantics communication. In *Findings of the Association for Computational Linguistics: EMNLP*, pages 84–93, 2020.

[Fei *et al.*, 2020c] Hao Fei, Yafeng Ren, and Donghong Ji. Mimic and conquer: Heterogeneous tree structure distillation for syntactic NLP. In *Findings of the Association for Computational Linguistics: EMNLP*, pages 183–193, 2020.

[Fei *et al.*, 2020d] Hao Fei, Yafeng Ren, and Donghong Ji. Retrofitting structure-aware transformer language model for end tasks. In *EMNLP*, pages 2151–2161, 2020.

[Fei *et al.*, 2021a] Hao Fei, Fei Li, Bobo Li, and Donghong Ji. Encoder-decoder based unified semantic role labeling with label-aware syntax. In *AAAI*, pages 12794–12802, 2021.

[Fei *et al.*, 2021b] Hao Fei, Shengqiong Wu, Yafeng Ren, Fei Li, and Donghong Ji. Better combine them together! integrating syntactic constituency and dependency representations for semantic role labeling. In *Findings of the Association for Computational Linguistics: ACL/IJCNLP 2021*, pages 549–559, 2021.

[Jia *et al.*, 2020] Qi Jia, Yizhu Liu, Siyu Ren, Kenny Zhu, and Haifeng Tang. Multi-turn response selection using dialogue dependency relations. In *EMNLP*, pages 1911–1920, 2020.

[Katiyar and Cardie, 2016] Arzoo Katiyar and Claire Cardie. Investigating LSTMs for joint extraction of opinion entities and relations. In *ACL*, pages 919–929, 2016.

[Long *et al.*, 2021] Xinwei Long, Shuzi Niu, and Yucheng Li. Consistent inference for dialogue relation extraction. In Zhi-Hua Zhou, editor, *IJCAI*, pages 3885–3891, 2021.

[Lu *et al.*, 2019] Junyu Lu, Chenbin Zhang, Zeying Xie, Guang Ling, Tom Chao Zhou, and Zenglin Xu. Constructing interpretive spatio-temporal features for multi-turn responses selection. In *ACL*, pages 44–50, 2019.

[Miwa and Bansal, 2016] Makoto Miwa and Mohit Bansal. End-to-end relation extraction using LSTMs on sequences and tree structures. In *ACL*, pages 1105–1116, 2016.

[Nan *et al.*, 2021] Guoshun Nan, Guoqing Luo, Sicong Leng, Yao Xiao, and Wei Lu. Speaker-oriented latent structures for dialogue-based relation extraction. *CoRR*, abs/2109.05182, 2021.

[Qiu *et al.*, 2021] Liang Qiu, Yuan Liang, Yizhou Zhao, Pan Lu, Baolin Peng, Zhou Yu, Ying Nian Wu, and Song-Chun Zhu. SocAoG: Incremental graph parsing for social relation inference in dialogues. In *ACL*, pages 658–670, 2021.

[Song *et al.*, 2019] Linfeng Song, Yue Zhang, Daniel Gildea, Mo Yu, Zhiguo Wang, and Jinsong Su. Leveraging dependency forest for neural medical relation extraction. In *EMNLP*, pages 208–218, 2019.

[Wu *et al.*, 2021] Shengqiong Wu, Hao Fei, Yafeng Ren, Donghong Ji, and Jingye Li. Learn from syntax: Improving pair-wise aspect and opinion terms extraction with rich syntactic knowledge. In *IJCAI*, pages 3957–3963, 2021.

[Xiong *et al.*, 2020] Ruibin Xiong, Yunchang Yang, Di He, Kai Zheng, Shuxin Zheng, Chen Xing, Huishuai Zhang, Yanyan Lan, Liwei Wang, and Tie-Yan Liu. On layer normalization in the transformer architecture. In *ICML*, pages 10524–10533, 2020.

[Xu *et al.*, 2015] Yan Xu, Lili Mou, Ge Li, Yunchuan Chen, Hao Peng, and Zhi Jin. Classifying relations via long short term memory networks along shortest dependency paths. In *EMNLP*, pages 1785–1794, 2015.

[Xue *et al.*, 2021] Fuzhao Xue, Aixin Sun, Hao Zhang, and Eng Siong Chng. Gdpnet: Refining latent multi-view graph for relation extraction. In *AAAI*, pages 14194–14202, 2021.

[Yao *et al.*, 2019] Yuan Yao, Deming Ye, Peng Li, Xu Han, Yankai Lin, Zhenghao Liu, Zhiyuan Liu, Lixin Huang, Jie Zhou, and Maosong Sun. DocRED: A large-scale document-level relation extraction dataset. In *ACL*, pages 764–777, 2019.

[Yu *et al.*, 2020] Dian Yu, Kai Sun, Claire Cardie, and Dong Yu. Dialogue-based relation extraction. In *ACL*, pages 4927–4940, 2020.

[Zhao *et al.*, 2021] Tianyang Zhao, Zhao Yan, Yunbo Cao, and Zhoujun Li. Enhancing dialogue-based relation extraction by speaker and trigger words prediction. In *Findings of ACL*, pages 4580–4585, 2021.

[Zhou *et al.*, 2021] Mengjia Zhou, Donghong Ji, and Fei Li. Relation extraction in dialogues: A deep learning model based on the generality and specialty of dialogue text. *IEEE ACM TASLP*, 29:2015–2026, 2021.