Speaker-Oriented Latent Structures for Dialogue-Based Relation Extraction

Guoshun Nan[1], **Guoqing Luo**[2*], **Sicong Leng**[1*], **Yao Xiao**[3] and **Wei Lu**[1]
[1]StatNLP Research Group, Singapore University of Technology and Design
[2]University of Alberta, Canada [3]Shanghai Jiao Tong University, China
nanguoshun@gmail.com, gluo@ualberta.ca
1190339100058@sjtu.edu.cn, sicong_leng@alumni.sutd.edu.sg
luwei@sutd.edu.sg

arXiv:2109.05182v1 [cs.CL] 11 Sep 2021

Abstract

Dialogue-based relation extraction (DiaRE) aims to detect the structural information from unstructured utterances in dialogues. Existing relation extraction models may be unsatisfactory under such a conversational setting, due to the *entangled logic* and *information sparsity* issues in utterances involving multiple speakers. To this end, we introduce SOLS, a novel model which can explicitly induce speaker-oriented latent structures for better DiaRE. Specifically, we learn latent structures to capture the relationships among tokens beyond the utterance boundaries, alleviating the entangled logic issue. During the learning process, our speaker-specific regularization method progressively highlights speaker-related key clues and erases the irrelevant ones, alleviating the information sparsity issue. Experiments on three public datasets demonstrate the effectiveness of our proposed approach.

1 Introduction

Relation extraction (RE) (Xu et al., 2015a,b; Peng et al., 2017) aims to detect structured relational information from unstructured texts. It has been widely used in various natural language processing (NLP) applications, such as knowledge graph construction (Huang et al., 2019; Zhang et al., 2021d), question answering (Xu et al., 2016; Zhu et al., 2021), and so on. Early studies towards this direction mainly focus on sentence-level RE (Zeng et al., 2014) that extracts information within a sentence. Recently, document-level RE (DRE) (Christopoulou et al., 2019; Yao et al., 2019; Jain et al., 2020) has drawn increasing attention as relationships between entities are often expressed across sentence boundaries.

Along the lines of DRE, a more interesting and challenging setting is dialogue-based RE (DiaRE)

S1: Jack. Could you come in here for a moment? Now!
S2: Found it.
S3: I'll take that **dad**.
S1: It seems your daughter and Richard are something of an item.
S2: That's impossible, he's got a twinkie in the city.
S4: Dad, I'm the twinkie.
S2: You're the twinkie?
S3: Yes, that is impossible
S5: She's not a twinkie.
... ...
S2: Am I supposed to stand here and listen to this on my birthday?
S4: **Dad, dad** this is a good thing for me. Ya know, and you even said yourself, you've never seen Richard happier.

Argument pair	Relation type
(S3, S4)	per: siblings
(S4, twinkie)	per: alternate_names
(S2, S4)	per: parents

Figure 1: An example adapted from the DialogRE dataset (Yu et al., 2020). In total, there are 5 speakers in the conversation covering different topics. S2, S3, and S4 indicate the abbreviations of different speakers.

(Yu et al., 2020; Zhang et al., 2020c), which aims to predict the relation between two arguments from a dialogue involving multiple speakers. Figure 1 illustrates a conversation snippet with 5 participants discussing several different topics. This instance is selected from the DialogRE (Yu et al., 2020) dataset collected from the scripts of the TV series "*Friends*". Here we give two examples from Figure 1 to demonstrate how the relations of entity pairs can be detected in dialogues.

1) To infer the relation for the pair ⟨S4, twinkie⟩, one needs to understand the utterance "Dad, I'm the twinkie" to identify it as *per:alternate_names*. Learning to predict such a relation is challenging due to entangled utterances such as "You're the twinkie?" and "She's not a twinkie", which may confuse a model when identifying the key clues.
2) To infer the relation between Speaker 3 (S3) and Speaker 4 (S4), we first need to identify the fact that Speaker 2 (S2) is the father of S3 from the utterance "I'll take that dad" at the beginning of the dialogue. Meanwhile,

*Equal contribution, work done during Guoqing Luo and Yao Xiao's internships at SUTD.

Accepted as a long paper in the main conference of EMNLP 2021 (Conference on Empirical Methods in Natural Language Processing).

we need to recognize that S4 also calls S2 "Dad" from the utterance "Dad, dad, this is . . . ". Combining such scattered information from each speaker, we can infer that the pair ⟨S3, S4⟩ forms a relation *per: siblings*.

From the above two cases, we observe that key to the success of DiaRE is capturing speaker-related information, which is also highlighted in the previous work DialogRE (Yu et al., 2020). Furthermore, most of the entity pairs in DiaRE are composed of one or two speakers: our statistics show that 89.9% of argument pairs involve at least one speaker in the DialogRE dataset. If we additionally consider the entities with *person* (PER) type as speakers, the above number can reach 99.9%. Although existing DRE methods have achieved great success, directly applying them to dialogues to detect such speaker-related contexts may be unsatisfactory, due to two underlying reasons listed as follows:

1) Conversations in DiaRE are often repetitive, and there may also be speaker interruptions (Sacks et al., 1978), leading to the *entangled logic* issue among utterances as shown in the first example, while a document in DRE is more narrative with much clearer logic.
2) There are many repetitive colloquial expressions such as "how are you?" and "yes" that are less informative for classification, resulting in the *information sparsity* issue (Yu et al., 2020) as shown in the second example.

Previous efforts have relied mainly on either a static graph (Chen et al., 2020c) or a matching mechanism (Zhang et al., 2020c) to aggregate the information for the DiaRE task. However, these heuristic rules may suffer from the above two issues under a dialogue setting, and may be unsatisfactory in handling the complex interactions among speakers and contexts. The recently proposed model GDPNet (Xue et al., 2021) improves rule-based methods by building multi-view latent graphs. However, the speaker-related context, which plays a crucial role for DiaRE, is not explicitly considered. Although pre-trained models have demonstrated effectiveness on RE tasks (Joshi et al., 2020; Wang et al., 2020), we will empirically show that these BERT-based models may still have limitations in capturing speaker-related contexts or addressing the two issues of DiaRE mentioned above.

For DiaRE, how to design a model that can effectively identify speaker-related context remains an open research problem. Inspired by GraphMask (Schlichtkrull et al., 2021) that learns to drop the unnecessary edges of a graph for model interpretation, we propose a novel method that explicitly induces *Speaker-Oriented Latent Structures* (SOLS) for better DiaRE. Experiments on three public datasets show the effectiveness of our SOLS approach. Our code and supplementary material are available at https://github.com/frankdarkluo/SOLS.

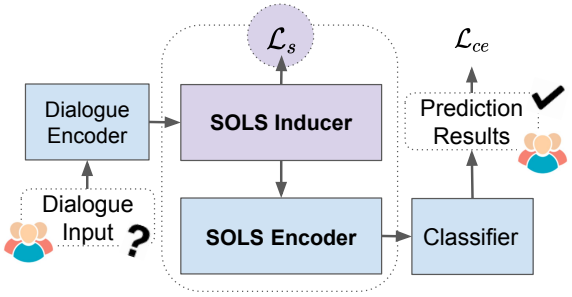


Figure 2: The architecture of our model.

The main contributions of this paper are:

- We propose a novel structure induction method to generate a latent graph for each argument (speaker) in dialogues, alleviating the entangled logic issue of DiaRE.
- We further introduce a novel speaker-oriented regularization method that explicitly highlights speaker-related salient contexts while discarding irrelevant ones, effectively addressing the information sparsity issue.
- We conduct quantitative and qualitative experiments on several public datasets to show the effectiveness of our approach, demonstrating the importance of capturing the speaker-related information in dialogues.

2 Model

2.1 Overview

Figure 2 shows the high-level overview of our model, which aims to obtain better DiaRE by exploring speaker-oriented latent structures. Our proposed model has four modules:

1) The dialogue encoder takes a dialogue as input and outputs contextualized representations.
2) The contextualized representations will be fed to our SOLS inducer to automatically generate two speaker-oriented latent structures with a novel regularization $\mathcal{L}_s$, aiming to mitigate the entangled logic and data sparsity issues. This is the core of our method.
3) The latent structures are then fed to SOLS encoder, which is a graph convolution network (GCN) (Kipf and Welling, 2017) for information aggregation.

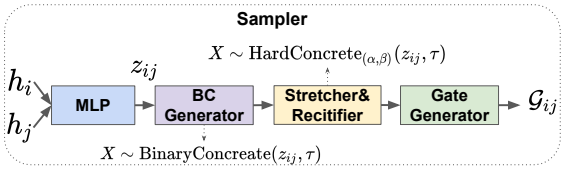


Figure 3: Our sampler, which learns a continuous gate $\mathcal{G}_{ij}$ for the representations of the i-th and j-th tokens. Such a gate is close to 0 or 1, representing the latent dependencies between two tokens in dialogues.

4) Finally, the classifier makes predictions.

Next, we detail how each module works.

2.2 Dialogue Encoder

We denote a dialogue as $\mathbf{d} = [x_1, \ldots, x_n]$ with n tokens and m utterances $[U_1, \ldots, U_m]$, where x_i is the i-th token and U_j is the j-th utterance in the dialogue. We treat $\mathbf{d}$ as a long sequence (n tokens) and feed it to a dialogue encoder, such as BiLSTM (Schuster and Paliwal, 1997), or a pre-trained BERT-based model (Devlin et al., 2019), to generate the contextualized representations $\mathbf{H} = [\mathbf{h}_1, ..., \mathbf{h}_n]$, where $\mathbf{h}_i \in \mathbb{R}^d$ is the contextualized representation of the i-th token with a dimension of d. Next we show how the latent structures can be induced based on the above representations.

2.3 SOLS Inducer

The intuition of SOLS is to learn a latent dialogue structure that is able to find out the speaker-related contexts, while ignoring the ones that are not or less relevant. Unlike previous studies that use the structured attention (Liu and Lapata, 2018; Isonuma et al., 2019; Nan et al., 2020) or Gaussian graph generator (Xue et al., 2021) to construct latent graphs, we employ the discrete-continuous distribution (Louizos et al., 2018) to explicitly learn speaker-oriented dialogue structures by sampling edge scores close to 0 or 1. We therefore regard a edge score in the graph as a "gate" and the score near 1 or 0 indicate turning on or off the connection between two tokens, respectively. Intuitively, the score value close to 1 indicates a strong relationship between two tokens and otherwise. We denote $\mathcal{G}_{ij} \in \mathbb{R}$ as the "gate" for the i-th and j-th tokens, and it can be computed by:

$$\mathcal{G}_{ij} = \mathcal{T}_\theta(\mathbf{h}_i, \mathbf{h}_j) \tag{1}$$

where $\mathcal{T}_\theta : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ is the gate sampler parameterized by θ, and $\mathbf{h}_i$ and $\mathbf{h}_j$ are the contextualized representations of the i-th and j-th token. Next we detail how our proposed sampler generates the "gate" $\mathcal{G}_{ij}$.

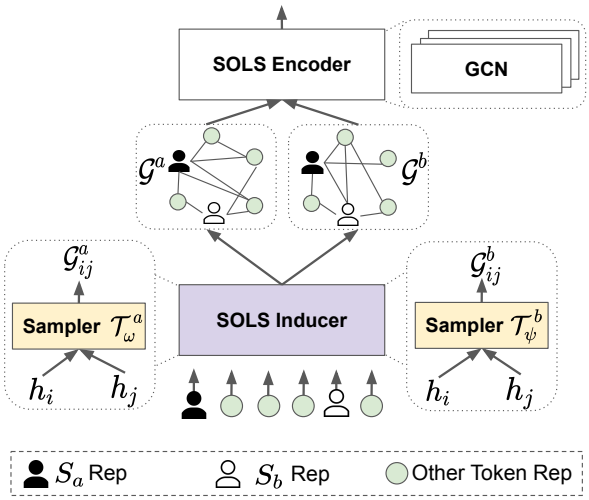


Figure 4: The proposed SOLS inducer and encoder. The inducer computes the dependency score between each two nodes in the dialogue. These learnable scores form two different latent graphs $\mathcal{G}^a$ and $\mathcal{G}^b$ for Speaker a and Speaker b, respectively. The two graphs will be fed into the SOLS encoder for information aggregation.

2.3.1 Sampling a Gate $\mathcal{G}_{ij}$

As shown in Figure 3, our sampler consists of four modules including MLP module, distribution generator, stretcher & rectifier, and gate generator.

1) MLP: For each i-j token pair, the MLP module takes their representations as input, and performs a non-linear transformation $\mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$, and outputs the scalar value $z_{ij} \in \mathbb{R}$ as:

$$z_{ij} = \text{MLP}([\mathbf{h}_i; \mathbf{h}_j]) \tag{2}$$

The above scalar z_{ij} is then regarded as a "learnable parameter", which will be used in the next component.

2) Distribution Generator: The distribution generator constructs a Binary Concrete (BC) distribution (Maddison et al., 2017) $X \sim \texttt{BinaryConcrete}(z_{ij}, \tau)$, using the learnable parameter z_{ij} outputted by the MLP module in Equation (2) and a fixed parameter τ. The BC distribution is composed of continuous discrete random variables based on the Gumbel-Max trick (Maddison, 2016). We use z_{ij} to control the probability mass skewing the BC distribution towards 0 or towards 1, in case of negative and positive locations respectively. Sampling values from such a distribution is analogous to generating a "gate" that can turn on or off the connection between two tokens.

3) Stretcher & Rectifier: As the BC random variables, which are generated in the previous step,

are defined over the open interval $(0, 1)$, the values 0 and 1 are unable to be sampled. Therefore, we further rely on the Hard Concrete (HC) distribution (Louizos et al., 2018) to extend the sampling from the open interval to the closed interval $[0, 1]$, by stretching and then rectifying the BC distribution using two parameters (α, β). For more details, interested readers are encouraged to read the Equation (10) and Equation (11) in Section 2.2 of the work Louizos et al. (2018). Specifically, HC distribution collapses the probability mass over the interval $(\alpha, 0]$ to 0, and the mass over the interval $[1, \beta)$ to 1, allowing the model to "highlight" the key context with a score close to 1 and "erase" some irrelevant ones with score close to 0.

4) Gate Generator: Finally, we sample a score $\mathcal{G}_{ij}$ for the i-th and j-th token from an HC distribution with the learnable parameter z_{ij} and the fixed parameter τ on the closed interval $[0, 1]$:

$$\begin{aligned} s_{ij} &= \sigma((\log \mu - \log(1-\mu) + z_{ij})/\tau) \\ \mathcal{G}_{ij} &= \min(1, \max(0, s_{ij} \times (\alpha - \beta) + \alpha)) \end{aligned} \tag{3}$$

where σ is the sigmoid function and $\mu \sim \mathcal{U}(0, 1)$ is sampled from a uniform distribution. By doing so, the latent dependencies between each pair of tokens in a dialogue can be learned with the above procedure.

2.3.2 Inducing Speaker-Oriented Structures

1) Latent Structure: Following the above procedure, we sample a gate for each pair of nodes in the dialogue $\mathbf{d}$ to construct a graph $\mathcal{G} \in \mathbb{R}^{n \times n}$. For a target relation, we generate two different graphs for each argument (speaker). Intuitively, each graph will place emphasis on its speaker-specific latent dependencies among tokens beyond the utterance boundaries. As such, the entangled logic issue, which is one of the main challenges of DiaRE highlighted at the beginning, could be properly alleviated with the latent relationships learned from the data. Figure 4 shows how we generate two structures $\mathcal{G}^a \in \mathbb{R}^{n \times n}$ and $\mathcal{G}^b \in \mathbb{R}^{n \times n}$ respectively for two speakers S_a and S_b, which are two arguments of the relation $\langle \mathrm{S}_a, \mathrm{S}_b \rangle$. The two graphs are expressed as follows:

$$\mathcal{G}^a := \left\{ \{ \mathcal{T}^a_\omega(\mathbf{h}_i, \mathbf{h}_j) \}; i, j \in [1, n] \right\} \tag{4}$$

$$\mathcal{G}^b := \left\{ \{ \mathcal{T}^b_\psi(\mathbf{h}_i, \mathbf{h}_j) \}; i, j \in [1, n] \right\} \tag{5}$$

where $\mathcal{T}^a_\omega$ and $\mathcal{T}^b_\psi$ refer to two samplers for S_a and S_b, parameterized by ω and ψ, respectively. For the argument pairs that involve non-speaker entities, we use the same sampling mechanism to generate latent graphs, as predicting relations for these pairs may similarly also have the entangled logic issues.

2) Controlled Sparsity: We have induced two latent graphs for two speakers S_a and S_b, which can be used as the adjacency matrix of GCNs for information aggregation. However, directly feeding the two graphs to GCNs may introduce noise for relation classification, as many of the contexts in dialogues could be irrelevant to the relation classification task. To alleviate this issue, we minimize the number of context tokens to be selected by introducing a regularization loss $\mathcal{L}_s$ during the induction of $\mathcal{G}^a$ and $\mathcal{G}^b$ to highlight the key clues, while dropping irrelevant connections. $\mathcal{L}_s$ can be considered as a controlled sparsity mechanism which minimizes the number of non-zeros predicted in the two graphs in a fully differentiable manner. We define $\mathcal{L}_s$ as:

$$\mathcal{L}_s = \mathbf{1}(\mathcal{G}^a) + \mathbf{1}(\mathcal{G}^b) \tag{6}$$

where $\mathbf{1}(.)$ is the indicator function that returns 1 if the input is non-zero. We use $\mathcal{G}^a_{ai} \in \mathcal{G}^a$ to denote the dependency score for S_a and the i-th node in the graph $\mathcal{G}^a$, and this gate can be expressed as:

$$\mathcal{G}^a_{ai} = \mathcal{T}^a_\theta(\mathbf{h}_i, \mathbf{h}_a), i \in [1, n] \tag{7}$$

Similarly, the gate $\mathcal{G}^b_{bi} \in \mathcal{G}^b$ represents the dependency score for S_b and the i-th node in the graph $\mathcal{G}^b$, and it can be expressed as:

$$\mathcal{G}^b_{bi} = \mathcal{T}^b_\theta(\mathbf{h}_i, \mathbf{h}_b), i \in [1, n] \tag{8}$$

Equipped with $\mathcal{L}_s$ in Equation (6), we are able to encourage the model to select the minimal number of key context with gates close to 1, and erase those with gates close to 0 during the induction of $\mathcal{G}^a$ and $\mathcal{G}^b$. Hence the information sparsity issue could be properly alleviated.

2.4 SOLS Encoder

Given two adjacency matrices $\mathcal{G}^a$ and $\mathcal{G}^b$, we employ GCN as a graph encoder for information aggregation. The convolution computation for the i-th node at the l-th layer takes the representation $\hat{\mathbf{h}}^{l-1}_i \in \mathbb{R}^d$ from the previous layer as an input and outputs the updated representation $\hat{\mathbf{h}}^l_i \in \mathbb{R}^d$:

$$\hat{\mathbf{h}}^l_i = \sigma\left(\sum_{j=1}^{n} \mathcal{G}_{ij} \mathbf{W}^l \hat{\mathbf{h}}^{l-1}_i + \mathbf{b}^l \right) \tag{9}$$

where $\mathbf{W}^l$ and $\mathbf{b}^l$ are the weight matrix and bias vector of the l-th layer respectively. Here $\hat{\mathbf{h}}^0_i \in \mathbb{R}^d$

indicates the initial contextual representation $\mathbf{h}_i$ of the i-th node. We use the same GCN for two speakers, and obtain the updated contextualized dialogue representations $\hat{\mathbf{H}}^a \in \mathbb{R}^{n\times d}$ and $\hat{\mathbf{H}}^b \in \mathbb{R}^{n\times d}$ after aggregation.

$$\hat{\mathbf{H}}^{a} = \mathrm{GCN}(\mathbf{H}, \mathcal{G}^{\mathrm{a}}), \quad \hat{\mathbf{H}}^{\mathrm{b}} = \mathrm{GCN}(\mathbf{H}, \mathcal{G}^{\mathrm{b}}) \tag{10}$$

2.5 Classifier and Loss Function

Finally, we use an MLP module as the classifier to predict the relation $r_{a,b}$ for the target argument pair $\langle \mathrm{S}_a, \mathrm{S}_b \rangle$:

$$r_{a,b} = \mathrm{MLP}([\hat{\mathbf{h}}^a; \hat{\mathbf{h}}^b]) \tag{11}$$

where $\hat{\mathbf{h}}_a \in \mathbb{R}^d$ and $\hat{\mathbf{h}}_b \in \mathbb{R}^d$ are the representations of the two speakers, generated by maxpooling over corresponding mention representations in $\hat{\mathbf{H}}^a$ and $\hat{\mathbf{H}}^b$. The overall loss $\mathcal{L}$ can be computed by:

$$\mathcal{L} = \mathcal{L}_{ce} + \lambda \mathcal{L}_s \tag{12}$$

where $\mathcal{L}_{ce}$ is the cross-entropy loss between the classification results and ground truth labels, and λ is the weight of speaker-related regularization.

3 Experiments

3.1 Datasets and Settings

We conduct experiments on three DiaRE datasets: (1) DialogRE English version (DialogRE-EN; Yu et al. 2020), the first human-annotated dialogue-level RE dataset from the famous American comedy "*Friends*"; (2) DialogRE Chinese version (DialogRE-CN; Yu et al. 2020), which is translated from DialogRE-EN; and (3) Medical Information Extractor (MIE; Zhang et al. 2020c), which involves doctor-patient dialogues collected from a Chinese medical consultation website. The statistics of three datasets are summarized in Table 1. We use the same data splits as the previous studies on existing three datasets. We leverage the Adam optimizer with learning rate 0.0001. The hidden size of the BiLSTM and GCNs are set as 200, and the number of layers of GCNs is configured as 1. The weight λ is set as 0.01. We refer to the previous work (Schlichtkrull et al., 2021) to fine-tune the Hard Concrete parameters, such as τ, α and β, which are set as 0.2, -0.2 and 1.2, respectively. The 300-dimensional GloVe (Pennington et al., 2014) is used to initialize the word embeddings for LSTM-based models[2].

[2]Details can be found in the supplementary material.

Dataset	Train	Dev	Test	Types
DialogRE-EN (Yu et al., 2020)	1,073	358	357	36
DiloagRE-CN (Yu et al., 2020)	1,073	358	357	36
MIE (Zhang et al., 2020c)	800	160	160	4

Table 1: Statistics of the three DiaRE datasets that are used in our experiments. We adapt the MIE dataset to our model following the setting of DialogRE.

3.2 Baselines

We compare our proposed SOLS method with various baselines, which are outlined as follows.

Sequence-based Models include some conventional neural networks such as CNN (Lawrence et al., 1997), LSTM (Schuster and Paliwal, 1997), and BiLSTM (Graves and Schmidhuber, 2005).

Rule-based Graph Models include C-GCN (Zhang et al., 2018) that constructs a graph with a pruned dependency tree, GCNN (Sahu et al., 2019) that relies on co-reference links to construct a document-level graph, EoG (Christopoulou et al., 2019) that uses pre-defined rules to build a document-level biomedical RE, and DHGAT (Chen et al., 2020c) that constructs a dialogue graph with a static rule and self-defined graph nodes.

Latent Graph Models involve AGGCN (Guo et al., 2019) that uses multi-head attention to build a graph, LSR (Nan et al., 2020) that employs Matrix-Tree Theorem (Koo et al., 2007) to generate and refine a document-level latent graph, and GDPNet (Xue et al., 2021) that uses Gaussian Graph Generator to build a multi-view latent graph.

BERT-based Models involve two popular pre-trained models BERT (base) (Devlin et al., 2019), RoBERTa (base) (Liu et al., 2019), as well as BERTs (Yu et al., 2020) under a conversational setting. We also compare the performance with SpanBERT (Joshi et al., 2020), which is a state-of-the-art pre-trained model for sentence-level RE. For the MIE dataset, we compare our model with previous attention-based MIE-Multi (Zhang et al., 2020c) method on BERT and RoBERTa.

3.3 Main Results

Table 2 summarizes the results on DialogRE-EN and DialogRE-CN datasets in terms of $F1$ and $F1_c$ scores. Here $F1_c$ is originally introduced by DialogRE (Yu et al., 2020), which considers the turns of dialogues when computing the $F1$. Comparisons and discussions are given as follows.

Comparisons with sequence-based models: We observe that the proposed SOLS method significantly outperforms baseline models CNN, LSTM

Type	Model	DialogRE-EN				DialogRE-CN			
		Dev		Test		Dev		Test	
		$F1$	$F1c$	$F1$	$F1c$	$F1$	$F1c$	$F1$	$F1c$
Sequential	CNN (Lawrence et al., 1997)	46.1*	43.7*	48.0*	45.0*	42.9	40.8	43.6	41.7
	LSTM (Schuster and Paliwal, 1997)	46.7*	44.2*	47.4*	44.9*	43.3	41.2	43.9	42.0
	BiLSTM (Graves and Schmidhuber, 2005)	48.1*	44.3*	48.6*	45.0*	44.4	41.7	44.8	42.3
Rule-based	C-GCN (Zhang et al., 2018)	45.8	40.1	44.3	40.3	40.2	39.3	40.5	39.7
	GCNN(Sahu et al., 2019)	47.3	44.2	48.2	45.1	44.1	41.5	44.3	42.1
	EoG(Christopoulou et al., 2019)	50.2	47.3	50.6	46.7	48.1	45.9	46.6	44.3
	DHGAT (BiLSTM) (Chen et al., 2020c)	57.7*	52.7*	56.1*	50.7*	55.8	53.6	54.6	52.7
Latent	AGGCN (Guo et al., 2019)	46.6*	40.5*	46.2*	39.5*	42.0	39.8	42.7	39.4
	LSR (BiLSTM) (Nan et al., 2020)	52.8	51.3	51.9	51.1	54.9	52.7	55.7	53.4
	GDPNet (BiLSTM) (Xue et al., 2021)	53.4	51.5	52.7	50.9	56.1	53.1	54.8	52.5
	Ours (BiLSTM)	**59.6**	**54.0**	**57.8**	**52.1**	**59.0**	**55.3**	**56.9**	**54.6**
BERT	BERT (Devlin et al., 2019)	60.6*	55.4*	58.5*	53.2*	63.7*	59.5*	63.2*	58.4*
	BERTs (Yu et al., 2020)	63.0*	57.3*	61.2*	55.4*	65.5*	61.0*	63.5*	58.7*
	RoBERTa (Liu et al., 2019)	65.2	61.4	62.8	58.8	64.0	59.8	62.7	58.9
	SpanBERT (Joshi et al., 2020)	64.6	58.8	61.8	55.8	–	–	–	–
	DHGAT (Chen et al., 2020c)	60.2	57.1	59.9	55.8	61.2	57.6	61.4	58.1
	LSR (Nan et al., 2020)	62.8	58.7	61.4	56.2	64.0	59.4	63.1	58.1
	GDPNet (Xue et al., 2021)	67.1*	61.5*	64.9*	60.1*	64.1	60.4	62.8	59.8
	Ours (BERT)	**69.6**	**62.6**	**68.1**	**61.4**	**66.7**	**61.6**	**65.4**	**60.6**

Table 2: Main results on DialogRE-EN and DialogRE-CN datasets. The results with * are directly taken from DialogRE (Yu et al., 2020), DHGAT (Chen et al., 2020c), or GDPNet (Xue et al., 2021). All other results are produced by us based on their open implementations as there are no previous results for these settings.

Model	P	R	F1
BERT (Devlin et al., 2019)	69.9	72.6	71.2
MIE-Multi (BERT) (Zhang et al., 2020c)	72.1	70.8	71.4
Ours (BERT)	**74.2**	**72.1**	**73.1**
RoBERTa (Liu et al., 2019)	70.6	70.9	70.7
MIE-Multi (RoBERTa) (Zhang et al., 2020c)	71.7	70.5	71.1
Ours (RoBERTa)	**72.6**	**71.9**	**72.2**

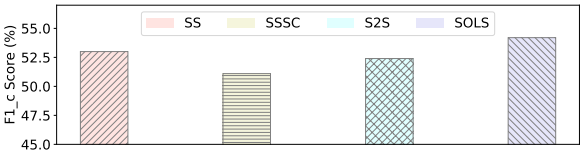
Table 3: Comparisons on the MIE dataset.

and BiLSTM by more than 9 points in terms of $F1$ and $F1_c$ on the development set of both datasets. Under the same BiLSTM encoder, SOLS achieves 7.1 and 12.5 points' improvement on the test set of both datasets in terms of $F1_c$, indicating the superiority of our model for DiaRE. The lower results on these baselines are probably due to the entangled logic and sparse information in dialogues, which makes it is difficult for sequential models to effectively encode the salient context.

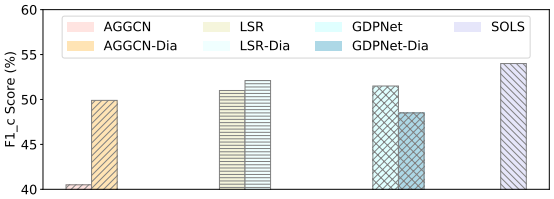
Comparisons with rule-based models: Compared with the rule-based baselines, our model with BiLSTM encoder consistently performs better on both datasets, giving more than 5 points improvement over C-GCN, GCNN and EoG in terms of $F1_c$. SOLS also outperforms a recent strong baseline DHGAT that builds an attention graph with a fixed aggregation path. The results show the capabilities of our model in learning more accurate relations between speakers and tokens in a multi-party dialogue. The "gates" learned by our model can dynamically turn on or off the connections between speakers and tokens with the mass of the value close to 0 or 1, whereas the above rule-based baselines use the pre-defined graph with prior knowledge, and hence are not flexible enough to handle the dialogue instances during the learning.

Comparisons with latent graph models: Compared with the latent graph baselines including AGGCN, LSR and GDPNet under the same BiLSTM encoder, our SOLS method is still able to yield the best performance, giving significant improvement on the two datasets, for example, 2.1 points improvement in terms of $F1$ against GDPNet on the test set of DialogRE-EN dataset. The comparisons confirm the effectiveness of our speaker-oriented latent structures in capturing the important clues for the dialogue-based relation extraction. Although the recent strong baseline GDPNet also induces a multi-view graph to capture various possible relationships among tokens, ignoring the modeling of speaker-oriented structures leads to much lower performance than our SOLS method.

Comparisons with BERT-based models: Our model consistently performs the best among all the BERT-based baselines on DialogRE-EN and DialogRE-CN datasets. Equipped with our SOLS method, we can improve BERT, BERTs, and DHGAT by 8.2, 6.0, and 5.6 points respectively in terms of $F1_c$ on the DialogRE-EN test set. Compared with SpanBERT, the state-of-the-art RE pre-trained model, we can still obtain much better scores. On the Chinese dataset DialogRE-CN, our model outperforms BERT-base and GDPNet

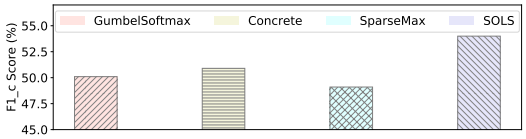


(a) Comparisons with SOLS variants.

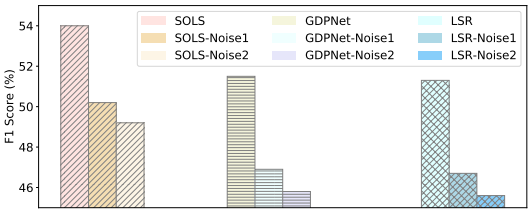


(b) Combining SOLS with latent graphs.

Figure 5: Discussions on the DialogRE-EN (dev set).



(a) Comparisons with various sparsity models.



(b) Comparisons under two different perturbations

Figure 6: Discussions on the DialogRE-EN (dev set).

by 2.2 and 2.6 points respectively in terms of $F1$ score. Note that the comparisons against SpanBERT on DialogRE-CN are not reported, since the pre-trained model is not available. The above results suggest that inducing speaker-oriented latent structures on a BERT-based model can further boost the performance, confirming our claim at the beginning that BERT-based models still have space for improvement for DiaRE. On the MIE dataset, Table 3 shows that SOLS obtains the best results under the same BERT-based encoders, further justifying the effectiveness of our latent structures.

3.4 Discussion

In this section, we discuss a few questions to better understand the latent structures learned by our proposed SOLS method.

Does the speaker-oriented learning paradigm matter? To answer this question, we introduce three new baselines that explicitly encourage the model to capture speaker-related contexts, including the variant SS of SOLS that highlights the connections among speakers generation of the gates, SS-SC that encourages the connections among speakers and the connections between speakers and tokens during graph generation, and S2S that encourages the connection between two speakers, and the tokens that appear in the text between them. Figure 5 (a) shows the comparison results on the development set of DialogRE-EN with the BiLSTM encoder. We observe that SS and S2S achieve 53.0 and 52.4 $F1_c$ and they perform better than AGGCN and LSR. The results suggest that introducing speaker-related constraint during structure induction will bring a performance gain.

Can we improve existing latent graph models with SOLS? To answer this question, we construct multi-view latent graphs by combing the latent structures induced by SOLS and some existing models, including AGGCN, LSR, and GDPNet. We name these new baselines as AGGCN-Dia, LSR-Dia, and GDPNet-Dia, each of which combines our latent graph to form a multi-view graph for information aggregation. Figure 5 (b) demonstrates comparisons between our model and these baselines on the development set of DialogRE-EN. In terms of $F1_c$, we observe that AGGCN-Dia and LSR-Dia obtain some improvements, indicating that our proposed method is able to further improve existing latent models by capturing dependencies between speakers and contexts. It is interesting to observe that GDPNet-Dia has the inferior performance than SOLS. This is probably due to too high complexity of combining two latent graphs in GDPNet and SOLS. We leave this interesting observation as our future research.

Can we highlight the key clues for extraction with existing sparsity methods? To answer this question, we leverage three sparsity techniques to output the sparse graph — GumbelSoftmax (Maddison, 2016), Concrete distribution (Maddison et al., 2017), and SparseMax (Martins and Astudillo, 2016). We use GDPNet to generate the adjacency matrix and then update the graph using GumbelSoftmax and SpareMax. We simply replace the Hard Concrete distribution in SOLS with the Concrete distribution[3]. Figure 6 (a) shows that our model consistently performs better than these baselines on the development set of DialogRE-EN, showing the superiority of the sampling method

[3]Parameters can be found in the supplementary material.

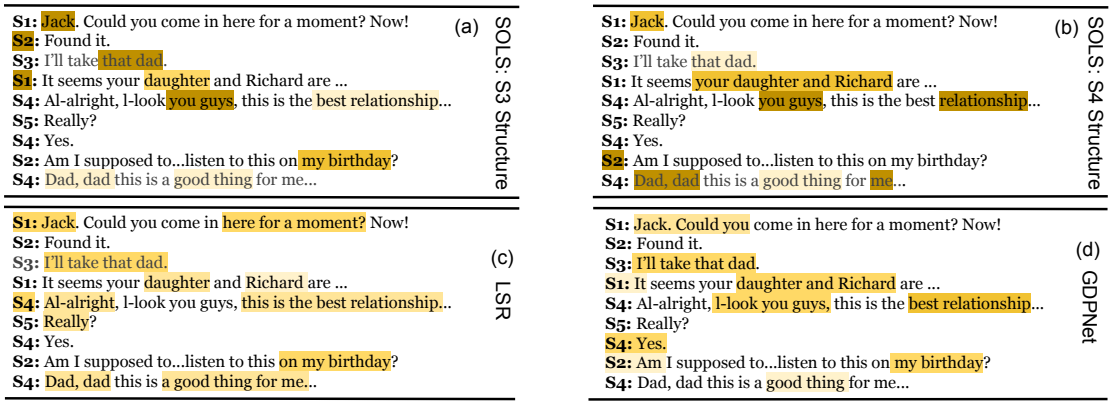


Figure 7: Case study on DialogRE-EN. The darker color means the higher score. Figure (a) shows the "gates" between Speaker 3 (S3) and the other tokens in the dialogue, and (b) shows the ones for Speaker 4 (S4). Figure (c) and (d) demonstrate the attention weights between S3 and the other tokens in the corresponding latent graphs.

Model	$F1$	$F1_c$
Full model	59.6	54.0
- SOLS	48.1	44.4
- Speaker-related regularization $\mathcal{L}_s$	54.6	52.3
- Graph $\mathcal{G}^a$	54.3	51.0
- Graph $\mathcal{G}^b$	56.1	52.7
- Gate	55.7	52.0
- Stretcher&Rectifier	56.0	52.4

Table 4: Ablation study on DialogRE-EN (deve set) with the BiLSTM encoder.

used in our SOLS method.

How robust is SOLS? Robustness is the key to the practical deployment of neural networks. To answer this question, we generate two noised datasets based on perturbations used in the previous work CLARE (Li et al., 2020) and TextAttack (Lütkebohle, 2021). The former substitutes a token with an alternative one based on RoBERTa, and the later randomly inserts some words in utterances[4]. We trained our method, LSR, and GDPNet on the clean DialogRE-EN training dataset, and then evaluate them on noised dev set. We add suffix "Noise1" or "Noise2" to indicate different perturbations, such as "SOLS-Noise1". Under the same setting, Figure 6 (b) shows that our SOLS method can yield more robust results compared with the two baselines, suggesting that the latent structures learned by our model can better resist perturbations.

3.5 Ablation Study

We conduct an ablation study on DialgRE-EN with the BERT encoder to evaluate the contribution of each component of our model. Table 4 reports the results. The removal of SOLS will lead to 11.5 and 9.6 points performance drops in terms of $F1$ and $F1_c$, respectively, quantitatively showing the gains obtained by our model over BiLSTM. Removing $\mathcal{L}_s$ decreases the $F1$ and $F1_c$ by 5.0 and 1.7 points, indicating that speaker-related regularization plays an important role for DiaRE. We also observe that removal of one latent structure lead to more than 3.5 $F1$ points performance drop on $F1$, suggesting that both speaker-oriented latent graphs contribute to the overall performance. The comparison also validates our hypothesis that each graph captures the salient context relevant to the corresponding speaker for relation classification. We observe that directly generating $\mathcal{G}_{ij}$ without three components including BC Generator, Stretcher&Rectifier and Gate Generator will decrease the scores by 3.9 points on $F1$. Bypassing the Stretcher&Rectifier module also leads to 3.6 points' drop. These results confirm the effectiveness of our proposed sampler.

3.6 Case Study

Figure 7 gives some qualitative analysis to see why our proposed model can better capture the speaker-related information for DiaRE. We visualize the latent graphs induced by SOLS, LSR and GDPNet using an instance selected from DialogRE-EN dataset with BERT encoder[5]. In this case, we aim to infer the relation between *Speaker 3* (S3) and *Speaker 4* (S4), which is indicated as *per:siblings*. As shown in Figure 7, the two latent graphs (a) and (b) induced for S3 and S4 are able to capture the key clues such as *Dad* with "gates" scores that are very close to 1, and thus the model can easily infer the relation of the two speakers by combining two latent graphs. However, LSR (c) and GDPNet

[4]Details can be found in the supplementary material.

[5]Whole dialogue is attached in the supplementary material.

(d) tend to yield more balanced attention between S3 and other tokens in the dialogue. Thus more irrelevant contexts will be selected for information aggregation, such as the tokens *"Really"* and *"Yes"*, leading to additional noise for relation classification.

4 Related Work

4.1 Relation Extraction

There have been many studies on sentence-level RE to detect relation facts in a given sentence (Zhou et al., 2020; Nguyen and Grishman, 2018; Song et al., 2018; Zhang et al., 2018; Lin et al., 2019; Wu and He, 2019; Guo et al., 2019; Jin et al., 2020; Zhang et al., 2020a; Guo et al., 2020; Zhang et al., 2020b; Qu et al., 2020; Zheng et al., 2021; Ye et al., 2021; Nan et al., 2021b), and document-level RE (Verga et al., 2018; Christopoulou et al., 2019; Yao et al., 2019; Jia et al., 2019; Jain et al., 2020; Xu et al., 2021b; Zhou et al., 2021; Wang et al., 2020; Zhang et al., 2021c) that try to detect structured knowledge from entire document. Compared to document-level RE, Dialogue-based relation extraction is more challenging, due to entangled logic and information sparsity issues, where the speaker-related information plays a critical role for relation predictions (Yu et al., 2020).

4.2 Dialogue-based Relation Extraction

DiaRE aims to extract interlocutor-related information in a dialogue. Previous work DialogRE (Yu et al., 2020) presents some sequence-based models such as CNN, BiLSTM (Schuster and Paliwal, 1997) for DiaRE, and proposes a BERTs model under the conversational setting as well. DHGAT (Chen et al., 2020c) introduces a predefined aggregation rule for DiaRE on graph attention networks (GAT). MIE (Zhang et al., 2020c) leverages a deep matching architecture to model doctor-patient dialogues. The above works rely on static rules for information aggregation and may not be flexible in handling the complex interactions among entities in dialogues. The recently proposed model GDPNet (Xue et al., 2021) constructs a latent multi-view graph to capture various potential relations among tokens. There are also some works for semantic representation for dialogue modeling based on AMR structure (Bai et al., 2021). Unlike these previous studies, we explicitly induce the speaker-oriented latent structures to capture complex interactions between speakers and contexts in conversations.

4.3 Latent Structure Induction

Latent structure models are powerful tools for composing information from contexts and building NLP pipelines (Martins et al., 2019). There are several approaches for the latent structure learning, including reinforcement learning (Havrylov et al., 2019), surrogate gradient (Corro and Titov, 2019a,b), and end-to-end differentiable methods (Kim et al., 2017; Isonuma et al., 2019; Chen et al., 2020a). SOLS falls into the third class that uses continuous relaxation to enable the gradient back-propagation, and is related to two recent works that try to interpret neural networks via the differentiable mask (De Cao et al., 2020; Schlichtkrull et al., 2021). Unlike these previous methods, SOLS learns the speaker-oriented latent structures with a novel speaker-specific regularization.

5 Conclusion and Future Work

This paper introduces SOLS, a novel model that aims to automatically induce the speaker-oriented latent structures for dialogue-based relation extraction. The model involves specifically designed modules based on the unique properties of such a dialogue-based task, where two issues, namely *entangled logic* and *information sparsity* issues are alleviated in the complex information extraction process. Experiments on three public DiaRE datasets show the effectiveness of our proposed SOLS method. In the future, we would like to generalize our model to more information extraction related tasks such as event detection and named entity recognition under a conversational setting. We are also interested in learning latent structures for visual and language tasks, such as video question answering (Xu et al., 2021a), visual dialogues (Fan et al., 2020), video grounding (Nan et al., 2021a; Liu et al., 2021; Zhang et al., 2021a,b), video relation detection (Li et al., 2021), and 3D captioning (Chen et al., 2020b, 2021).

Acknowledgments

We would like to thank the anonymous reviewers for their thoughtful and constructive comments. This research is supported by Ministry of Education, Singapore, under its Academic Research Fund (AcRF) Tier 2 Programme (MOE AcRF Tier 2 Award No: MOE2017-T2-1-156). Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not reflect the views of the Ministry of Education, Singapore.

References

Xuefeng Bai, Yulong Chen, Linfeng Song, and Yue Zhang. 2021. Semantic representation for dialogue modeling. In *Proc. of ACL*.

Chenhua Chen, Zhiyang Teng, and Yue Zhang. 2020a. Inducing target-specific latent structures for aspect sentiment classification. In *Proc. of EMNLP*.

Dave Zhenyu Chen, Angel X Chang, and Matthias Nießner. 2020b. Scanrefer: 3d object localization in rgb-d scans using natural language. In *Proc. of ECCV*.

Hui Chen, Pengfei Hong, Wei Han, Navonil Majumder, and Soujanya Poria. 2020c. Dialogue relation extraction with document-level heterogeneous graph attention networks. *arXiv preprint arXiv:2009.05092*.

Zhenyu Chen, Ali Gholami, Matthias Nießner, and Angel X Chang. 2021. Scan2cap: Context-aware dense captioning in rgb-d scans. In *Proc. of CVPR*.

Fenia Christopoulou, Makoto Miwa, and Sophia Ananiadou. 2019. Connecting the dots: Document-level neural relation extraction with edge-oriented graphs. In *Proc. of EMNLP*.

Caio Corro and Ivan Titov. 2019a. Differentiable perturb-and-parse: Semi-supervised parsing with a structured variational autoencoder. In *Proc. of ICLR*.

Caio Corro and Ivan Titov. 2019b. Learning latent trees with stochastic perturbations and differentiable dynamic programming. In *Proc. of ACL*.

Nicola De Cao, Michael Sejr Schlichtkrull, Wilker Aziz, and Ivan Titov. 2020. How do decisions emerge across layers in neural models? interpretation with differentiable masking. In *Proc. of EMNLP*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proc. of NAACL*.

Hehe Fan, Linchao Zhu, Yi Yang, and Fei Wu. 2020. Recurrent attention network with reinforced generator for visual dialog. *ACM Transactions on Multimedia Computing, Communications, and Applications TOMM*, 16(3):78:1–78:16.

Alex Graves and Jürgen Schmidhuber. 2005. Framewise phoneme classification with bidirectional lstm and other neural network architectures. *Neural networks*, 18(5-6):602–610.

Zhijiang Guo, Guoshun Nan, Wei Lu, and Shay B Cohen. 2020. Learning latent forests for medical relation extraction. In *Proc. of IJCAI*.

Zhijiang Guo, Yan Zhang, and Wei Lu. 2019. Attention guided graph convolutional networks for relation extraction. In *Proc. of ACL*.

Serhii Havrylov, Germán Kruszewski, and Armand Joulin. 2019. Cooperative learning of disjoint syntax and semantics. In *Proc. of NAACL-HLT*.

Xiao Huang, Jingyuan Zhang, Dingcheng Li, and Ping Li. 2019. Knowledge graph embedding based question answering. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, pages 105–113.

Masaru Isonuma, Junichiro Mori, and Ichiro Sakata. 2019. Unsupervised neural single-document summarization of reviews via learning latent discourse structure and its ranking. In *Proc. of ACL*.

Sarthak Jain, Madeleine van Zuylen, Hannaneh Hajishirzi, and Iz Beltagy. 2020. Scirex: A challenge dataset for document-level information extraction. In *Proc. of ACL*.

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

Lifeng Jin, Linfeng Song, Yue Zhang, Kun Xu, Weiyun Ma, and Dong Yu. 2020. Relation extraction exploiting full dependency forests. In *Proc. of AAAI*.

Mandar Joshi, Danqi Chen, Yinhan Liu, Daniel S Weld, Luke Zettlemoyer, and Omer Levy. 2020. Spanbert: Improving pre-training by representing and predicting spans. *Transactions of the Association for Computational Linguistics*, 8:64–77.

Yoon Kim, Carl Denton, Luong Hoang, and Alexander M. Rush. 2017. Structured attention networks. In *Proc. of ICLR*.

Thomas N. Kipf and Max Welling. 2017. Semi-supervised classification with graph convolutional networks. In *Proc. of ICLR*.

Terry K Koo, Amir Globerson, Xavier Carreras, and Michael Collins. 2007. Structured prediction models via the matrix-tree theorem. In *Proc. of EMNLP-CoNLL*.

Steve Lawrence, C Lee Giles, Ah Chung Tsoi, and Andrew D Back. 1997. Face recognition: A convolutional neural-network approach. *IEEE transactions on neural networks*, 8(1):98–113.

Dianqi Li, Yizhe Zhang, Hao Peng, Liqun Chen, Chris Brockett, Ming-Ting Sun, and Bill Dolan. 2020. Contextualized perturbation for textual adversarial attack. *arXiv preprint arXiv:2009.07502*.

Yicong Li, Xun Yang, Xindi Shang, and Tat-Seng Chua. 2021. Interventional video relation detection. In *Proc. of ACM MM*.

Hongtao Lin, Jun Yan, Meng Qu, and Xiang Ren. 2019. Learning dual retrieval module for semi-supervised relation extraction. In *Proc. of WWW*, pages 1073–1083.

Daizong Liu, Xiaoye Qu, Jianfeng Dong, Pan Zhou, Yu Cheng, Wei Wei, Zichuan Xu, and Yulai Xie. 2021. Context-aware biaffine localizing network for temporal sentence grounding. In *Proc. of CVPR*.

Yang Liu and Mirella Lapata. 2018. Learning structured text representations. *Transactions of the Association for Computational Linguistics*, 6:63–75.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Christos Louizos, Max Welling, and Diederik P Kingma. 2018. Learning sparse neural networks through l_0 regularization. In *Proc. of ICLR*.

Ingo Lütkebohle. 2021. BWorld Robot Control Software. https://textattack.readthedocs.io/en/latest/3recipes/models.html. [Online; accessed 15-May-2021].

CA Maddison. 2016. Poisson process model for monte carlo. *Perturbation, Optimization, and Statistics*, pages 193–232.

Chris J Maddison, Andriy Mnih, and Yee Whye Teh. 2017. The concrete distribution: A continuous relaxation of discrete random variables. In *Proc. of ICLR*.

Andre Martins and Ramon Astudillo. 2016. From softmax to sparsemax: A sparse model of attention and multi-label classification. In *Proc. of ICML*.

André F. T. Martins, Tsvetomila Mihaylova, Nikita Nangia, and Vlad Niculae. 2019. Latent structure models for natural language processing. In *Proc. of ACL*.

Guoshun Nan, Zhijiang Guo, Ivan Sekulić, and Wei Lu. 2020. Reasoning with latent structure refinement for document-level relation extraction. In *Proc. of ACL*.

Guoshun Nan, Rui Qiao, Yao Xiao, Jun Liu, Sicong Leng, Hao Zhang, and Wei Lu. 2021a. Interventional video grounding with dual contrastive learning. In *Proc. of CVPR*.

Guoshun Nan, Jiaqi Zeng, Rui Qiao, Zhijiang Guo, and Wei Lu. 2021b. Uncovering main causalities for long-tailed information extraction. In *Proc. of EMNLP*.

Thien Huu Nguyen and Ralph Grishman. 2018. Graph convolutional networks with argument-aware pooling for event detection. In *Proc. of AAAI*.

Nanyun Peng, Hoifung Poon, Chris Quirk, Kristina Toutanova, and Wen tau Yih. 2017. Cross-sentence n-ary relation extraction with graph lstms. *Transactions of the Association for Computational Linguistics*, 5:101–115.

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global vectors for word representation. In *Proc. of EMNLP*.

Meng Qu, Tianyu Gao, Louis-Pascal Xhonneux, and Jian Tang. 2020. Few-shot relation extraction via bayesian meta-learning on relation graphs. In *Proc. of ICML*.

Harvey Sacks, Emanuel A Schegloff, and Gail Jefferson. 1978. A simplest systematics for the organization of turn taking for conversation. In *Studies in the organization of conversational interaction*, pages 7–55. Elsevier.

Sunil Kumar Sahu, Fenia Christopoulou, Makoto Miwa, and Sophia Ananiadou. 2019. Inter-sentence relation extraction with document-level graph convolutional neural network. In *Proc. of ACL*.

Michael Sejr Schlichtkrull, Nicola De Cao, and Ivan Titov. 2021. Interpreting graph neural networks for nlp with differentiable edge masking. In *Proc. of ICLR*.

Mike Schuster and Kuldip K. Paliwal. 1997. Bidirectional recurrent neural networks. *IEEE Trans. Signal Processing*, 45:2673–2681.

Linfeng Song, Yue Zhang, Zhiguo Wang, and Daniel Gildea. 2018. N-ary relation extraction using graph-state lstm. In *Proc. of EMNLP*.

Patrick Verga, Emma Strubell, and Andrew McCallum. 2018. Simultaneously self-attending to all mentions for full-abstract biological relation extraction. In *Proc. of NAACL-HLT*.

Difeng Wang, Wei Hu, Ermei Cao, and Weijian Sun. 2020. Global-to-local neural networks for document-level relation extraction. In *Proc. of EMNLP*.

Shanchan Wu and Yifan He. 2019. Enriching pre-trained language model with entity information for relation classification. In *Proc. of CIKM*.

Kun Xu, Yansong Feng, Songfang Huang, and Dongyan Zhao. 2015a. Semantic relation classification via convolutional neural networks with simple negative sampling. In *Proc. of EMNLP*.

Kun Xu, Siva Reddy, Yansong Feng, Songfang Huang, and Dongyan Zhao. 2016. Question answering on freebase via relation extraction and textual evidence. In *Proc. of ACL*.

Li Xu, He Huang, and Jun Liu. 2021a. Sutd-trafficqa: A question answering benchmark and an efficient network for video reasoning over traffic events. In *Proc. of CVPR*.

Wang Xu, Kehai Chen, and Tiejun Zhao. 2021b. Document-level relation extraction with reconstruction. In *Proc. of AAAI*.

Yan Xu, Lili Mou, Ge Li, Yunchuan Chen, Hao Peng, and Zhi Jin. 2015b. Classifying relations via long short term memory networks along shortest dependency paths. In *Proc. of EMNLP*.

Fuzhao Xue, Aixin Sun, Hao Zhang, and Eng Siong Chng. 2021. Gdpnet: Refining latent multi-view graph for relation extraction. In *Proc. of AAAI*.

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 *Proc. of ACL*.

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

Dian Yu, Kai Sun, Claire Cardie, and Dong Yu. 2020. Dialogue-based relation extraction. In *Proc. of ACL*.

Daojian Zeng, Kang Liu, Siwei Lai, Guangyou Zhou, and Jian Zhao. 2014. Relation classification via convolutional deep neural network. In *Proc. of COLING*.

Hao Zhang, Aixin Sun, Wei Jing, Guoshun Nan, Liangli Zhen, Joey Tianyi Zhou, and Rick Siow Mong Goh. 2021a. Video corpus moment retrieval with contrastive learning. In *Proc. of SigIR*.

Hao Zhang, Aixin Sun, Wei Jing, Liangli Zhen, Joey Tianyi Zhou, and Rick Siow Mong Goh. 2021b. Natural language video localization: A revisit in span-based question answering framework. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Ningyu Zhang, Xiang Chen, Xin Xie, Shumin Deng, Chuanqi Tan, Mosha Chen, Fei Huang, Luo Si, and Huajun Chen. 2021c. Document-level relation extraction as semantic segmentation. In *Proc. of IJCAI*.

Ningyu Zhang, Shumin Deng, Zhen Bi, Haiyang Yu, Jiacheng Yang, Mosha Chen, Fei Huang, Wei Zhang, and Huajun Chen. 2020a. Openue: An open toolkit of universal extraction from text. In *Proc of. EMNLP*.

Ningyu Zhang, Shumin Deng, Zhanlin Sun, Jiaoyan Chen, Wei Zhang, and Huajun Chen. 2020b. Relation adversarial network for low resource knowledge graph completion. In *Proc. of WWW*.

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

Yuanzhe Zhang, Zhongtao Jiang, Tao Zhang, Shiwan Liu, Jiarun Cao, Kang Liu, Shengping Liu, and Jun Zhao. 2020c. MIE: A medical information extractor towards medical dialogues. In *Proc. of ACL*.

Yuhao Zhang, Peng Qi, and Christopher D Manning. 2018. Graph convolution over pruned dependency trees improves relation extraction. In *Proc. of EMNLP*.

Hengyi Zheng, Rui Wen, Xi Chen, Yifan Yang, Yunyan Zhang, Ziheng Zhang, Ningyu Zhang, Bin Qin, Ming Xu, and Yefeng Zheng. 2021. Prgc: Potential relation and global correspondence based joint relational triple extraction. In *Proc. of ACL*.

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

Wenxuan Zhou, Hongtao Lin, Bill Yuchen Lin, Ziqi Wang, Junyi Du, Leonardo Neves, and Xiang Ren. 2020. Nero: A neural rule grounding framework for label-efficient relation extraction. In *Proc. of WWW*.

Fengbin Zhu, Wenqiang Lei, Youcheng Huang, Chao Wang, Shuo Zhang, Jiancheng Lv, Fuli Feng, and Tat-Seng Chua. 2021. TAT-QA: A question answering benchmark on a hybrid of tabular and textual content in finance. In *Proc. of ACL*.