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 3Shanghai Jiao Tong University, China
nanguoshun@gmail.com, gluo@ualberta.ca
119033910058@sjtu.edu.cn, sicong_leng@alumni.sutd.edu.sg
luwei@sutd.edu.sg

——EMNLP2021

https://github.com/frankdarkluo/SOLS

Reported by Yabo Yin

1.Introduction

2.Method

3.Experiments

Introduction

1. Conversations in DiaRE are often repetitive,and there may also be speaker interruptions, leading to the entangled logic issue among utterances.
2. There are many repetitive colloquial expressions to lead less informative for classification, resulting in the information sparsity issue.

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.

Method

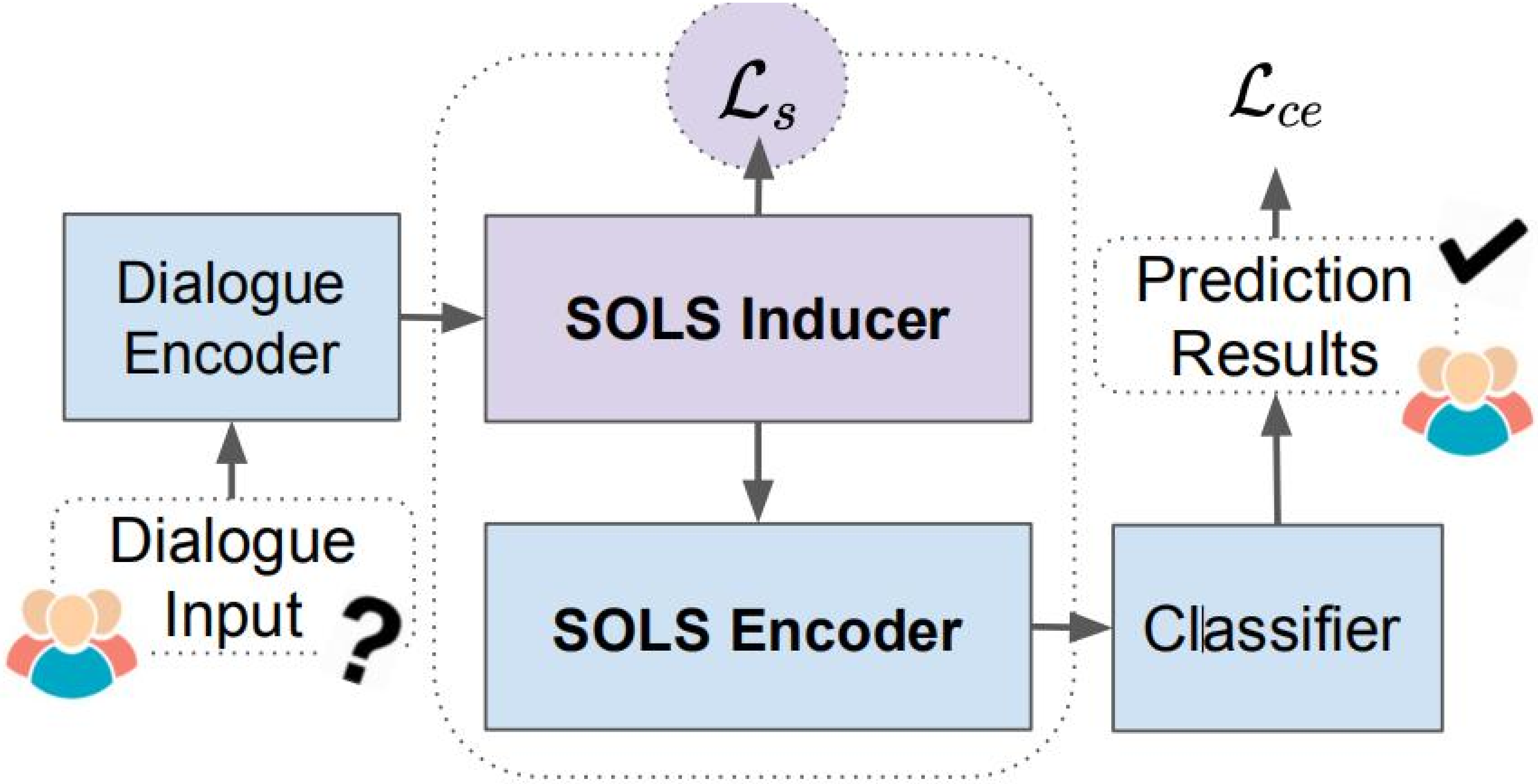


Figure 2: The architecture of our model.

Method

Dialogue Encoder

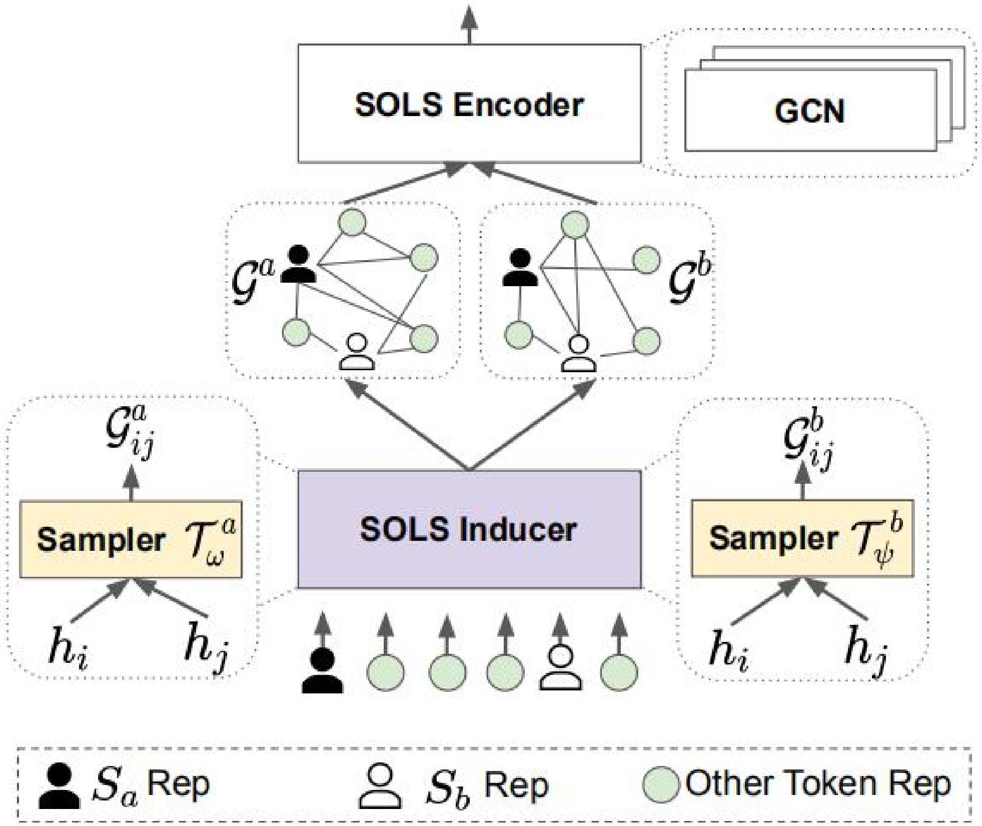
a dialogue as $\mathbf{d} = [x_1, \ldots, x_n]$

m utterances $[U_1, \ldots, U_m]$

contextualized representations $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_n]$ $\mathbf{h}_i \in \mathbb{R}^d$

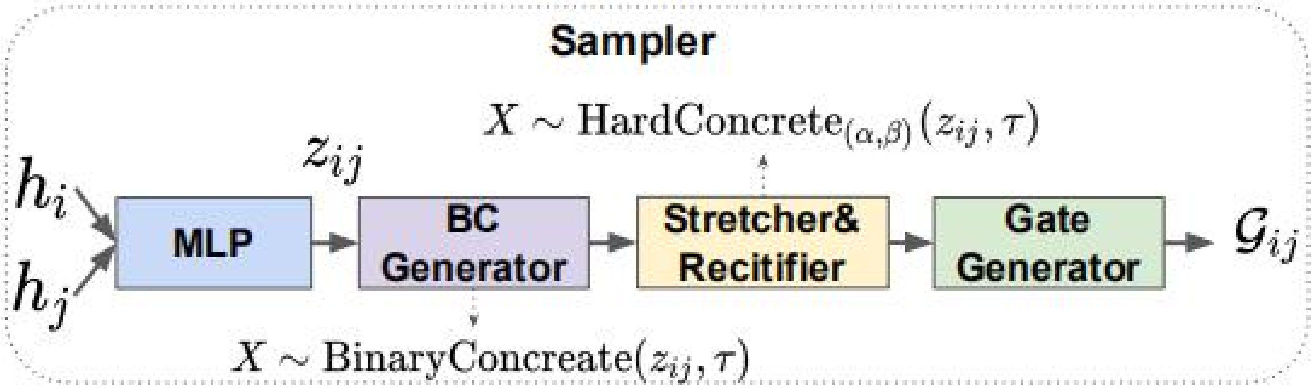
$$\mathcal{G}_{ij} = \mathcal{T}_\theta(\mathbf{h}_i, \mathbf{h}_j) \qquad (1)$$

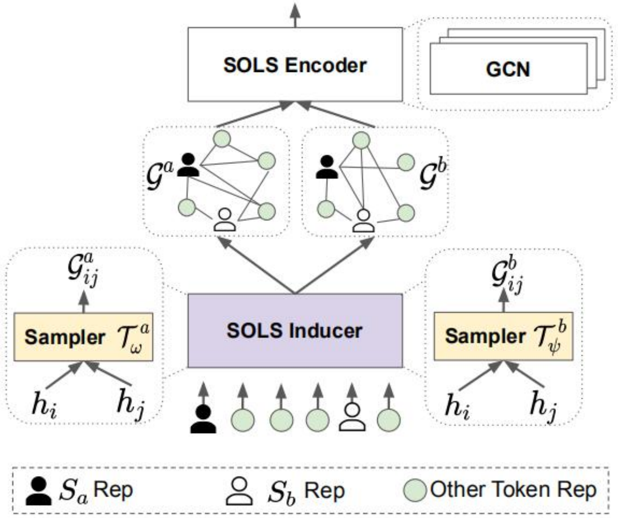
where $\mathcal{T}_\theta : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ is the gate sampler parameterized by θ, and $\mathbf{h}_i$ and $\mathbf{h}_j$ are the contextualized representations of the i-th and j-th token. Next



Method

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





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

$$X \sim \texttt{BinaryConcrete}(z_{ij}, \tau)$$ 约等于z/(1+z)

$$X \sim \textbf{HardConcrete}_{(\alpha,\beta)}(z_{ij}, \tau)$$ 将(0,1)=>[0,1]

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

where σ is the sigmoid function and $\mu \sim \mathcal{U}(0,1)$ is sampled from a uniform distribution.

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

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

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

Method

Controlled Sparsity:

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

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

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

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

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

$$r_{a,b} = \mathrm{MLP}([\hat{\mathbf{h}}^a; \hat{\mathbf{h}}^b]) \tag{11}$$ $\hat{\mathbf{h}}_a \in \mathbb{R}^d$ and $\hat{\mathbf{h}}_b \in \mathbb{R}^d$

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

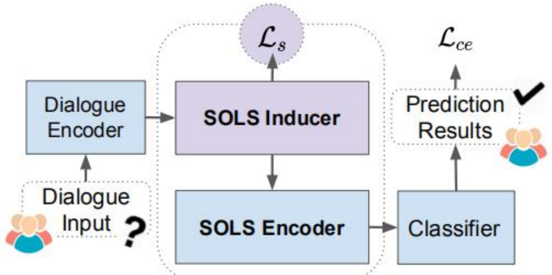
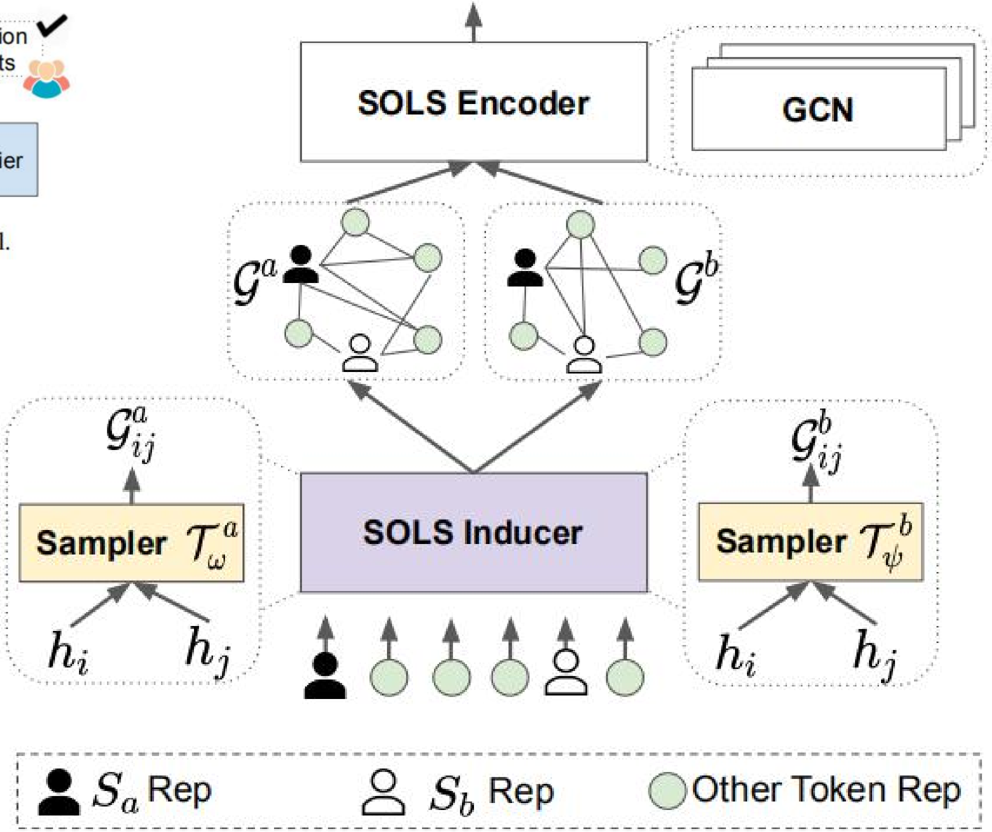


Figure 2: The architecture of our model.



Experiments

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.

Experiments

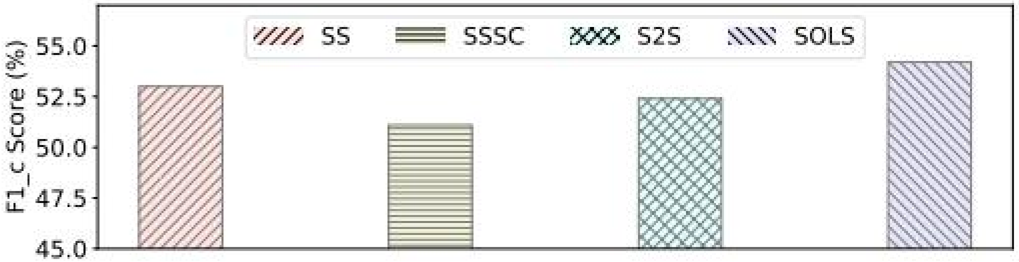
Model	P	R	F1
BERT (Devlin et al., 2019)	69.9	72.6	71.2
MIE-Multi (BERT) (Zhang et al., 2020c)	72.1	70.8	71.4
Ours (BERT)	**74.2**	**72.1**	**73.1**
RoBERTa (Liu et al., 2019)	70.6	70.9	70.7
MIE-Multi (RoBERTa) (Zhang et al., 2020c)	71.7	70.5	71.1
Ours (RoBERTa)	**72.6**	**71.9**	**72.2**

Table 3: Comparisons on the MIE dataset.

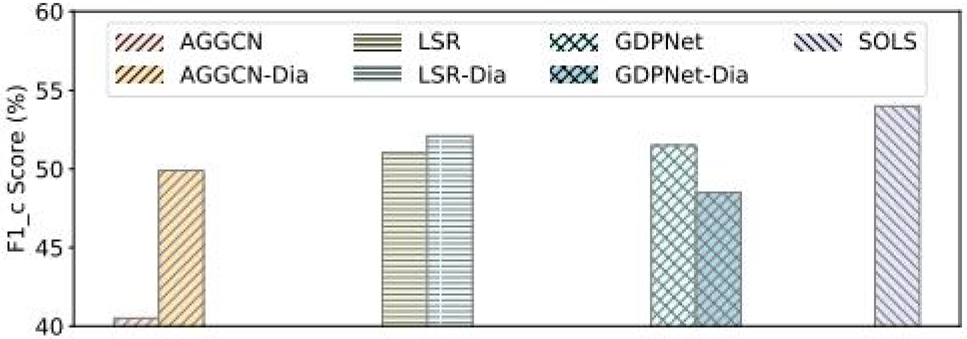
Model	$F1$	$F1_c$
Full model	59.6	54.0
- SOLS	48.1	44.4
- Speaker-related regularization $\mathcal{L}_s$	54.6	52.3
- Graph $\mathcal{G}^a$	54.3	51.0
- Graph $\mathcal{G}^b$	56.1	52.7
- Gate	55.7	52.0
- Stretcher&Rectifier	56.0	52.4

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

Experiments

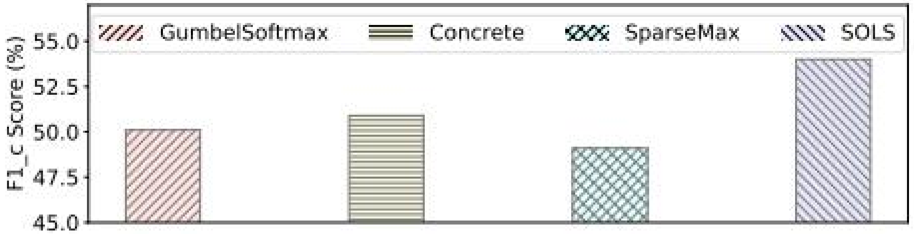


(a) Comparisons with SOLS variants.

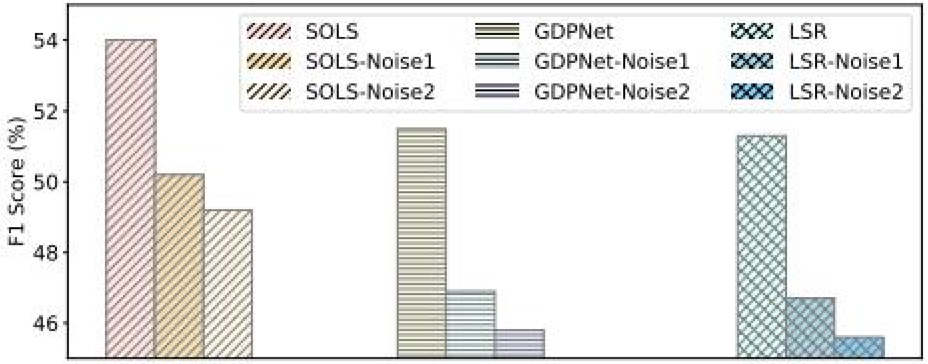


(b) Combining SOLS with latent graphs.

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).

Experiments

(a) SOLS: S3 Structure

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 ...
S4: Al-alright, l-look you guys, this is the best relationship...
S5: Really?
S4: Yes.
S2: Am I supposed to...listen to this on my birthday?
S4: Dad, dad this is a good thing for me...

(b) SOLS: S4 Structure

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 ...
S4: Al-alright, l-look you guys, this is the best relationship...
S5: Really?
S4: Yes.
S2: Am I supposed to...listen to this on my birthday?
S4: Dad, dad this is a good thing for me...

(c) LSR

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 ...
S4: Al-alright, l-look you guys, this is the best relationship...
S5: Really?
S4: Yes.
S2: Am I supposed to...listen to this on my birthday?
S4: Dad, dad this is a good thing for me...

(d) GDPNet

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 ...
S4: Al-alright, l-look you guys, this is the best relationship...
S5: Really?
S4: Yes.
S2: Am I supposed to...listen to this on my birthday?
S4: Dad, dad this is a good thing for me...

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.

Thank you!