Dynamic Schema Graph Fusion Network for Multi-Domain Dialogue State Tracking

Yue Feng[†] Aldo Lipani[†] Fanghua Ye[†] Qiang Zhang[‡] * Emine Yilmaz[†]
[†]University College London, London, UK
[‡]Zhejiang University, Hangzhou, China
[†] {yue.feng.20,aldo.lipani,fanghua.ye.19,emine.yilmaz}@ucl.ac.uk
[‡]qiang.zhang.cs@zju.edu.cn

Code：https://github.com/sweetalyssum/DSGFNet.

（ACL-2022)

gesis
Leibniz-Institut für Sozialwissenschaften

Reported by Zicong Dou

1. Introduction
2. Approach
3. Experiments

Introduction

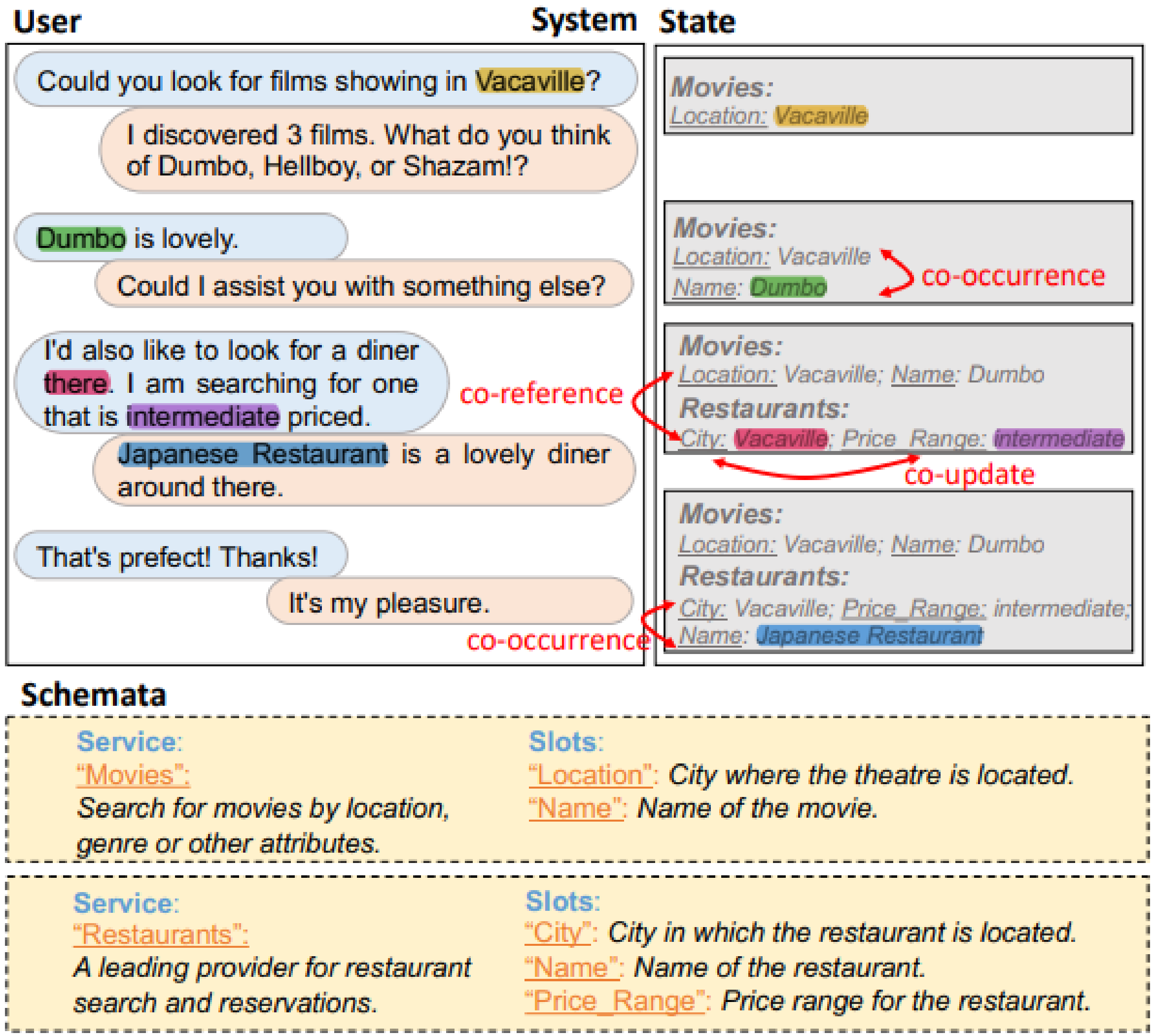


Figure 1: An example of DST. Given the schemata for all domains, the slot values are extracted from the user and system utterances (e.g., spans highlighted with the same color in the figure). The dialogue state of each turn is represented as a set of slot-value pairs. Among the domains and slots, there are prior slot-domain membership relations which are expressed in the predefined schemata, and also dialogue-aware dynamic slot relations which depend on the dialogue context (e.g., co-reference, co-update, and co-occurrence).

Co-reference：当一个槽值在对话中早先被提及并被分配给另一个槽时，就会发生共指关系；
Co-update：当槽值在同一个对话回合中一起更新时，会发生协同更新关系
Co-occurrence：在大型对话语料库中同现概率高的槽在当前对话中一起出现时。

Approach

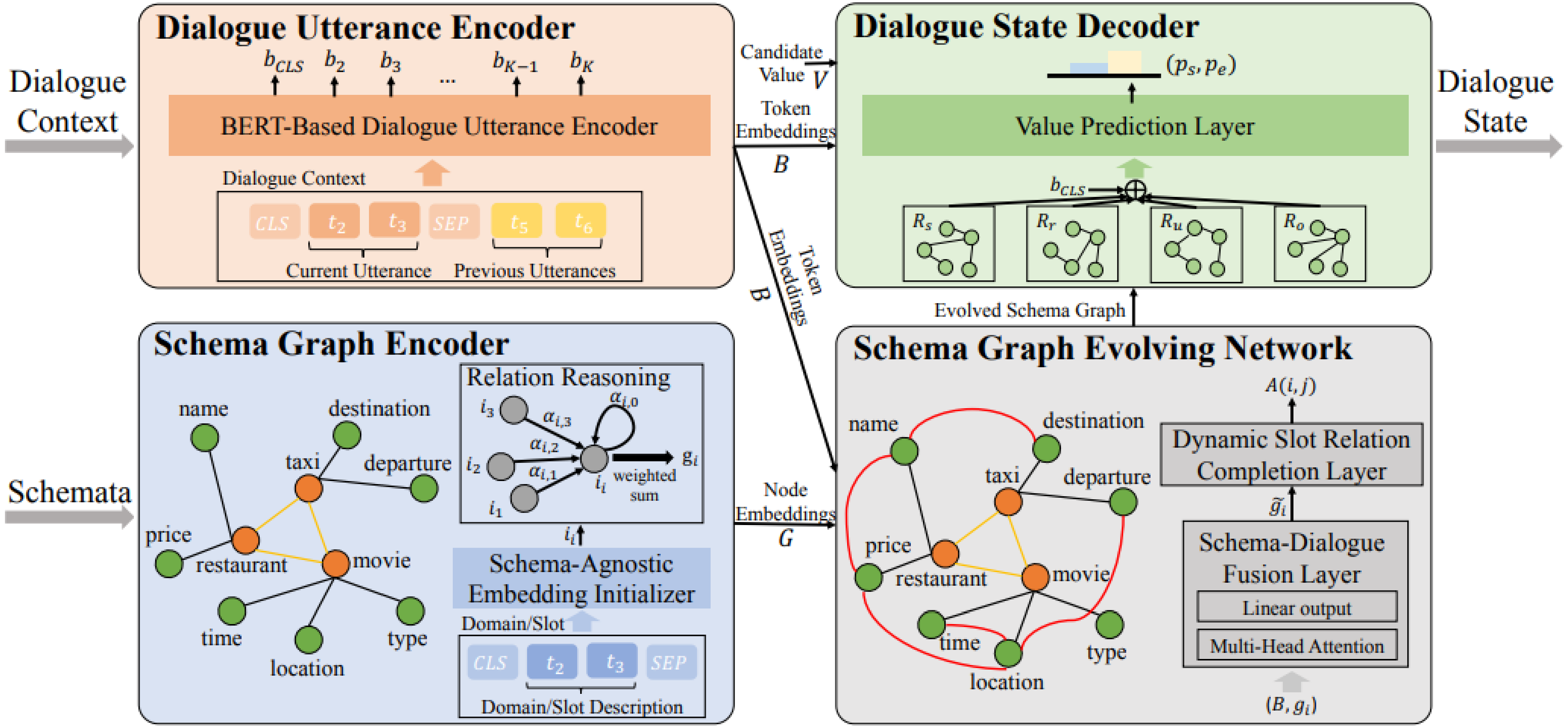
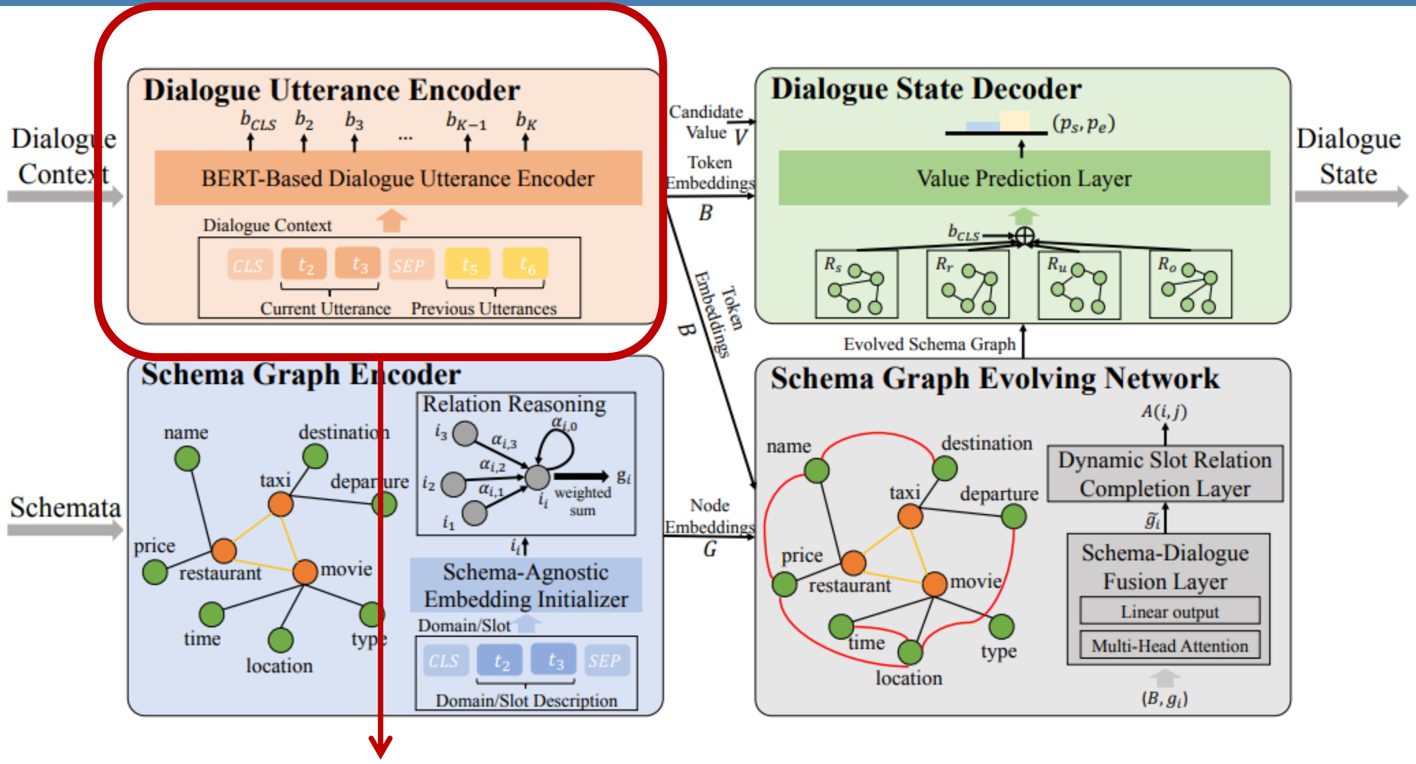
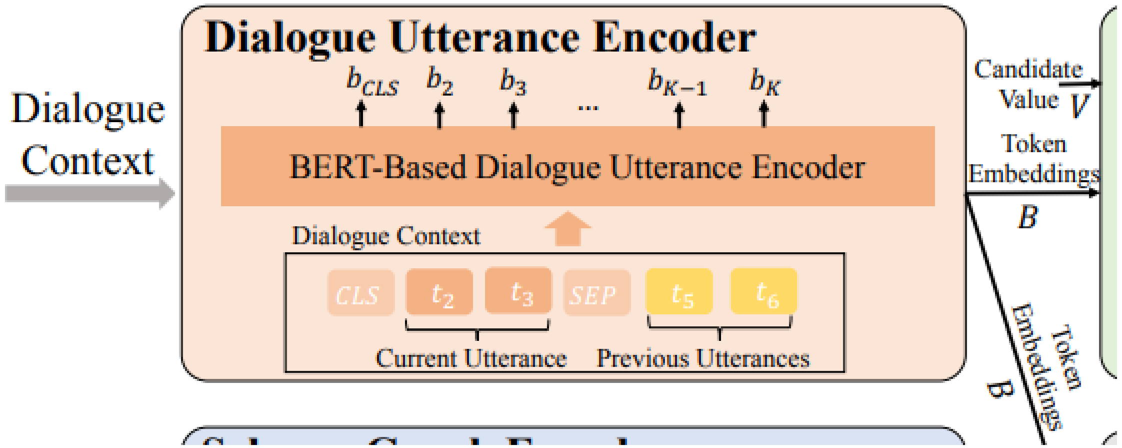


Figure 2: The architecture of DSGFNet, which contains a dialogue utterance encoder, a schema graph encoder, a schema graph evolving network, and a dialogue state decoder.





Dialogue Utterance Encoder

$$[t_1, \ldots, t_K]$$

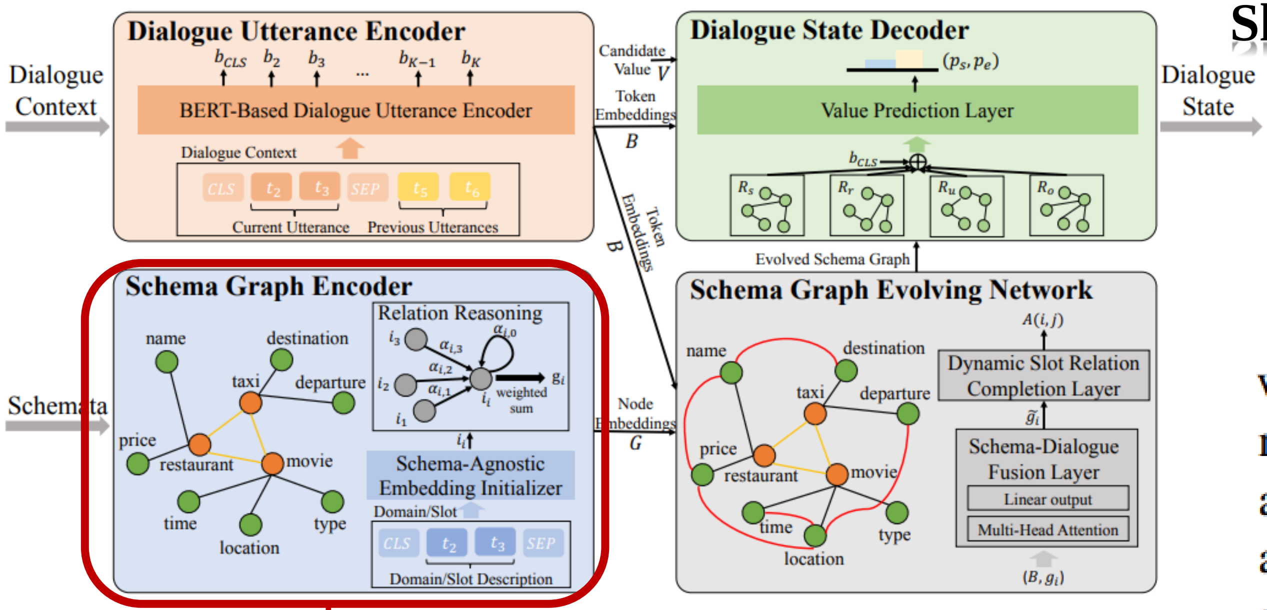
$$B = [b_1, \ldots, b_K]$$

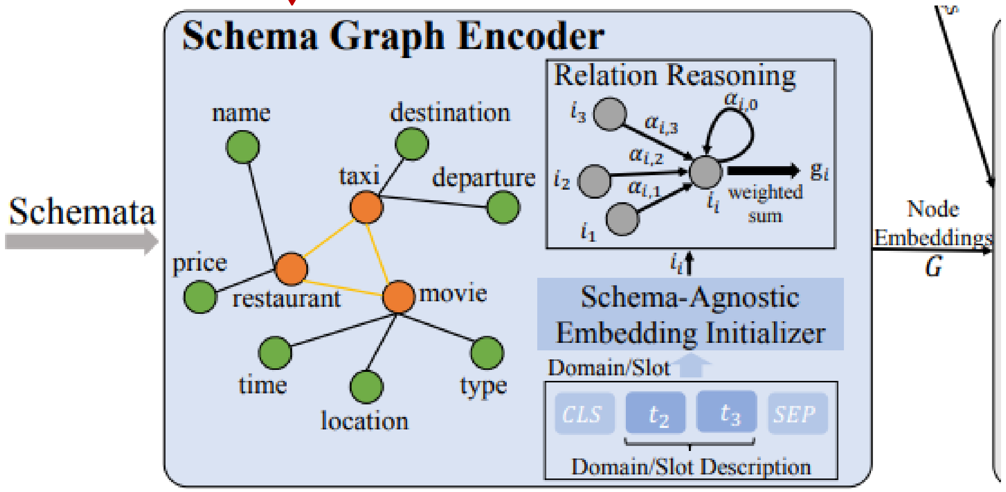
Schema Graph Encoder

Schema-Agnostic Embedding Initializer

$$I = [i_1, \ldots, i_{N+M}]$$

Approach





Slot-Domain Membership Relation Reasoning Network

$$h_{i,j} = \mathrm{ReLU}(\mathbf{W}^\top \cdot [\boldsymbol{i}_i, \boldsymbol{i}_j]), \tag{1}$$

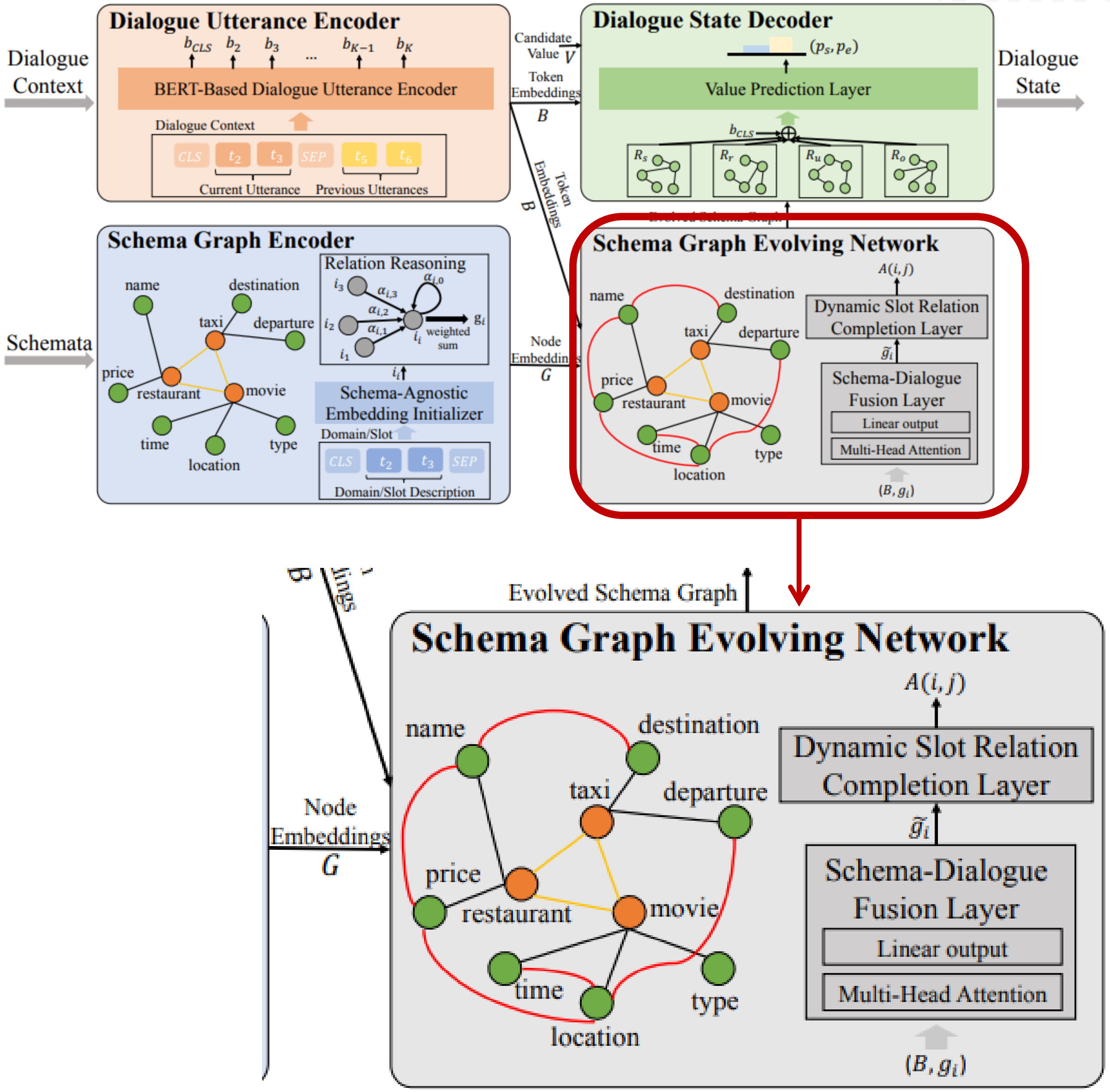
$$\alpha_{i,j} = \frac{\exp(h_{i,j})}{\sum_{k \in \mathcal{N}_i} \exp(h_{i,k})}, \tag{2}$$

where $\mathbf{W}$ is a matrix of parameters and $\mathcal{N}_i$ is the neighborhood of the i-th node. The normalized attention coefficients and the activation function are used to compute a non-linear weighted combination of the neighbours. This is used to compute the tensor of the schema graph node embeddings $\boldsymbol{G} = (\boldsymbol{g}_1, ..., \boldsymbol{g}_{N+M})$:

$$\boldsymbol{g}_i = \mathrm{ReLU}\left(\sum_{j \in \mathcal{N}_i} \alpha_{i,j} \cdot \boldsymbol{i}_j\right), \tag{3}$$

where $i \in \{1, \ldots, N+M\}$. To explore the higher-order connectivity information of slots across domains, we stack l layers of the reasoning network.

Approach



Schema Graph Evolving Network

Schema-Dialogue Fusion Layer

$$\boldsymbol{H} = \mathrm{MultiHead}(\mathrm{Q} = \boldsymbol{g}_i, \mathrm{K} = \boldsymbol{B}, \mathrm{V} = \boldsymbol{B}), \quad (4)$$

$$\tilde{\boldsymbol{g}}_i = \boldsymbol{H} \cdot \boldsymbol{W}_a, \quad (5)$$

where $\boldsymbol{W}_a$ is learnable parameters of a linear projection after the multi-head attention, and $\tilde{\boldsymbol{g}}_i$ is the dialogue-aware schema graph node embeddings.

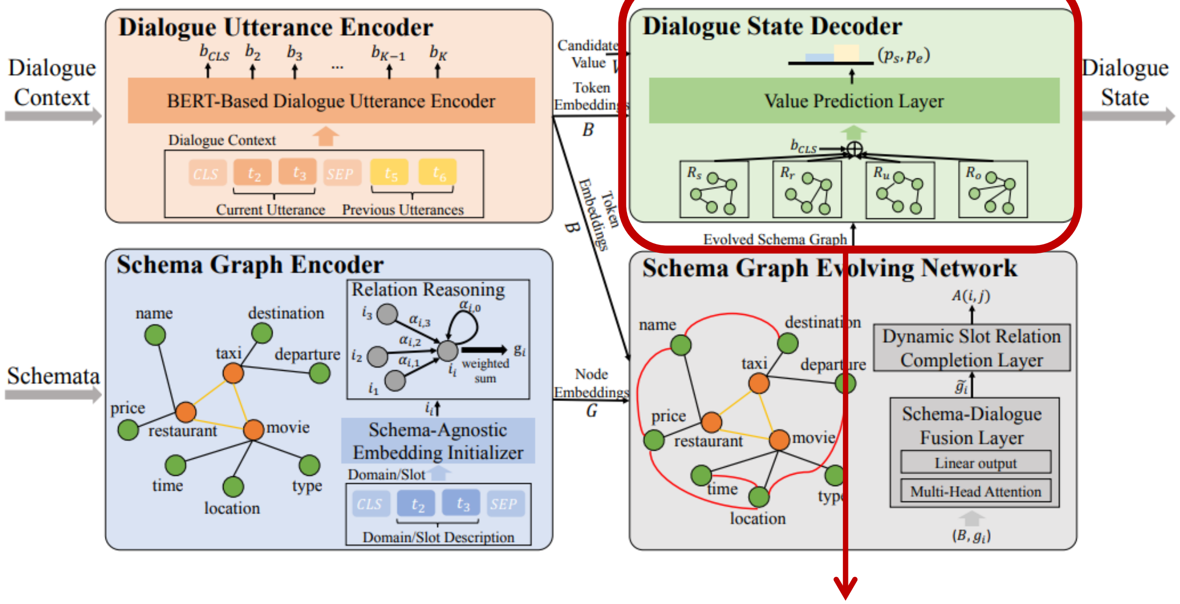
Dynamic Slot Relation Completion Layer

$$\boldsymbol{A}(i, j) = \arg\max\left(\mathrm{softmax}(\mathrm{MLP}(\tilde{\boldsymbol{g}}_i \oplus \tilde{\boldsymbol{g}}_j))\right). \quad (6)$$

above. Formally, given the i-th and j-th dialogue-aware slot node embeddings $\tilde{\boldsymbol{g}}_i$ and $\tilde{\boldsymbol{g}}_j$, we obtain an adjacent matrix of the dynamic slot relations for all slot pairs as follows:

With $\boldsymbol{A}$, we add dynamic slot relation edges to the schema graph.

Approach



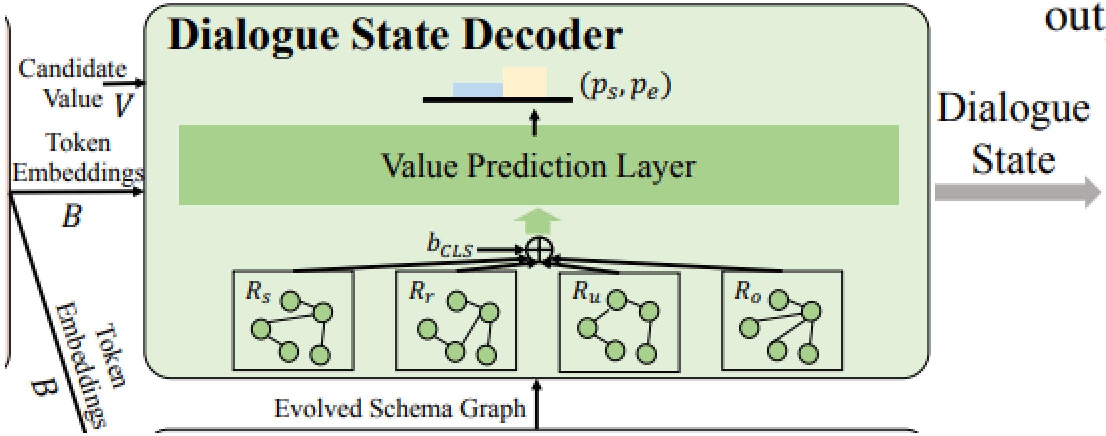
Dialogue State Decoder

$$\boldsymbol{S} = [\boldsymbol{s}_s; \boldsymbol{s}_r; \boldsymbol{s}_u; \boldsymbol{s}_o], \tag{7}$$

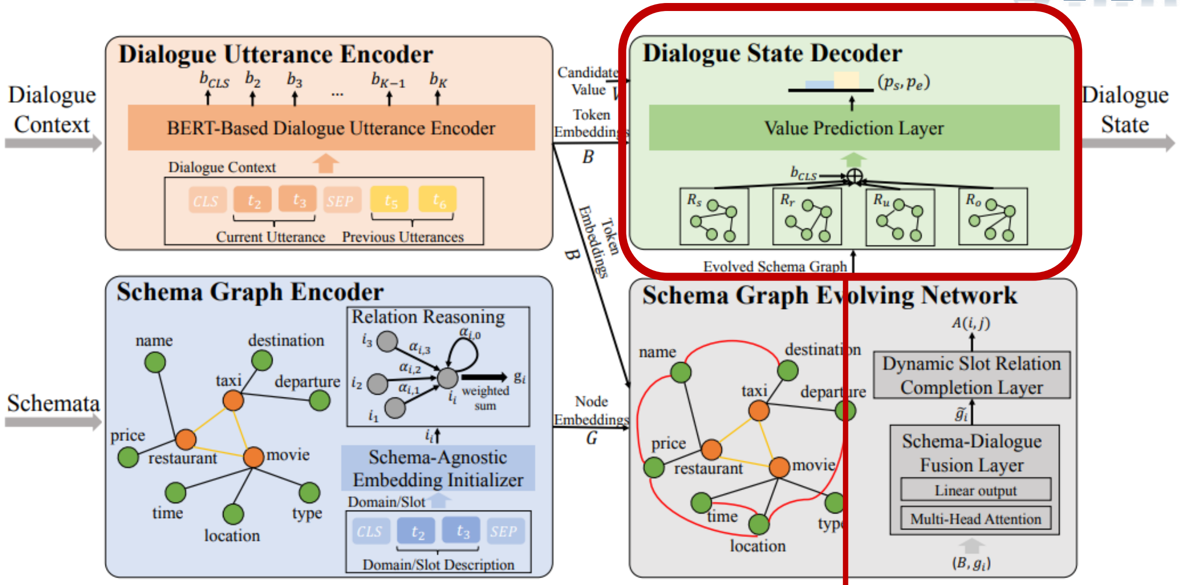
$$\boldsymbol{\beta} = \mathrm{softmax}(\boldsymbol{S}^\top \cdot \tanh(\boldsymbol{W}_s \cdot \boldsymbol{b}_{[CLS]} + \boldsymbol{b}_s)), \tag{8}$$

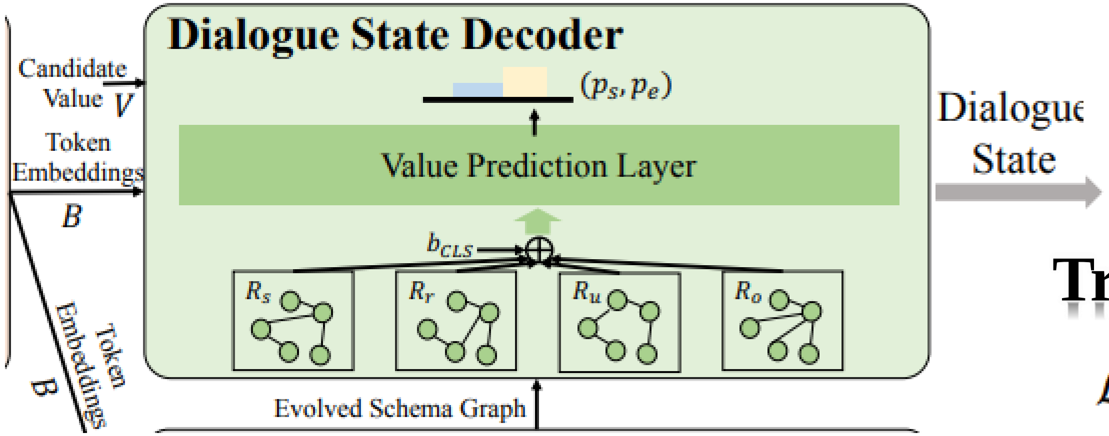
$$\boldsymbol{s} = \boldsymbol{S} \cdot \boldsymbol{\beta}, \tag{9}$$

where $\boldsymbol{W}_s$, $\boldsymbol{b}_s$ are learnable weights, $\boldsymbol{b}_{[CLS]}$ is the output of BERT-based dialogue utterance encoder.



Approach





Dialogue State Decoder

Each slot value is extracted by a value predictor based on the corresponding fused slot node embeddings $\boldsymbol{s}$. The value predictor is a trainable nonlinear classifier followed by two parallel softmax layers to predict start and end positions in candidate elements $\boldsymbol{C}$, which are composed by the dialogue context $\boldsymbol{B}$ and slots' candidate value vocabulary $\boldsymbol{V}$:

$$\boldsymbol{C} = [\boldsymbol{B}; \boldsymbol{V}] \tag{10}$$

$$[\boldsymbol{l}_s, \boldsymbol{l}_e] = \boldsymbol{r}_d \cdot \tanh(\boldsymbol{s}^\top \cdot \boldsymbol{W}_d \cdot \boldsymbol{C} + \boldsymbol{b}_d), \tag{11}$$

$$p_s = \mathrm{softmax}(\boldsymbol{l}_s), \tag{12}$$

$$p_e = \mathrm{softmax}(\boldsymbol{l}_e), \tag{13}$$

where $\boldsymbol{r}_d$, $\boldsymbol{W}_d$, and $\boldsymbol{b}_d$ are trainable parameters.

Training and Inference

$$\mathcal{L} = \lambda \cdot \mathcal{L}_r + (1 - \lambda) \cdot \mathcal{L}_s, \tag{14}$$

where $\lambda \in [0, 1]$ is a balance coefficient.

Experiments

Table 1: Characteristics of the datasets in experiments. The numbers provided are for the training sets of the corresponding datasets.

Characteristics	**SGD**	**MultiWOZ2.2**	**MultiWOZ2.1**
No. of domains	16	8	7
No. of dialogues	16,142	8,438	8,438
Total no. of turns	329,964	113,556	113,556
Avg. turns per dialogue	20.44	13.46	13.46
Avg. tokens per turn	9.75	13.13	13.38
No. of slots	215	61	37
Unseen domains in test set	Yes	No	No

Table 2: Joint GA of DSGFNet and baselines in unseen domains and all domains on SGD dataset. DSGFNet significantly improves over the best baseline (two-sided paired t-test, $p < 0.05$).

Models	**SGD Unseen Domains**	**SGD All Domains**
SGD-baseline (Rastogi et al., 2020)	20.0%	25.4%
FastSGT (Noroozi et al., 2020)	20.8%	29.2%
Seq2Seq-DU (Feng et al., 2021)	23.5%	30.1%
DSGFNet	**24.4%**	**32.1%**

Experiments

Table 3: Joint GA of DSGFNet and baselines on MultiWOZ2.2. DSGFNet significantly improves over the best baseline (two-sided paired t-test, $p < 0.05$).

Model	MultiWOZ2.2
SGD-baseline (Rastogi et al., 2020)	42.0%
TRADE (Wu et al., 2019)	45.4%
DS-DST (Zhang et al., 2020a)	51.7%
TripPy (Heck et al., 2020)	53.5%
Seq2Seq-DU (Feng et al., 2021)	54.4%
DSGFNet	**55.8%**

Table 4: Joint GA of DSGFNet and baselines on MultiWOZ2.1. DSGFNet achieves comparable performance of the best baseline.

Model	MultiWOZ2.1
SGD-baseline (Rastogi et al., 2020)	43.4%
TRADE (Wu et al., 2019)	46.0%
DS-DST (Zhang et al., 2020a)	51.2%
SOM-DST (Kim et al., 2020)	53.0%
MinTL-BART (Lin et al., 2020)	53.6%
SST (Chen et al., 2020)	55.2%
TripPy (Heck et al., 2020)	55.3%
PPTOD (Su et al., 2021)	**57.1%**
DSGFNet	56.7%

Experiments

Table 5: Ablation study on unseen domains of SGD, all domains of SGD, MultiWOZ2.2 and MultiWOZ2.1.

Model	Joint GA Unseen Domains SGD	Joint GA All Domains SGD	Joint GA MultiWOZ 2.2	Joint GA MultiWOZ 2.1
DSGFNet	24.4%	32.1%	55.8%	56.7%
-w/o Slot-Domain Membership Relations	21.9%	29.8%	53.4%	54.1%
-w/o Dynamic Slot Relations	20.6%	28.6%	52.2%	53.2%
-w/o Relation Aggregation	23.8%	31.5%	55.2%	55.9%

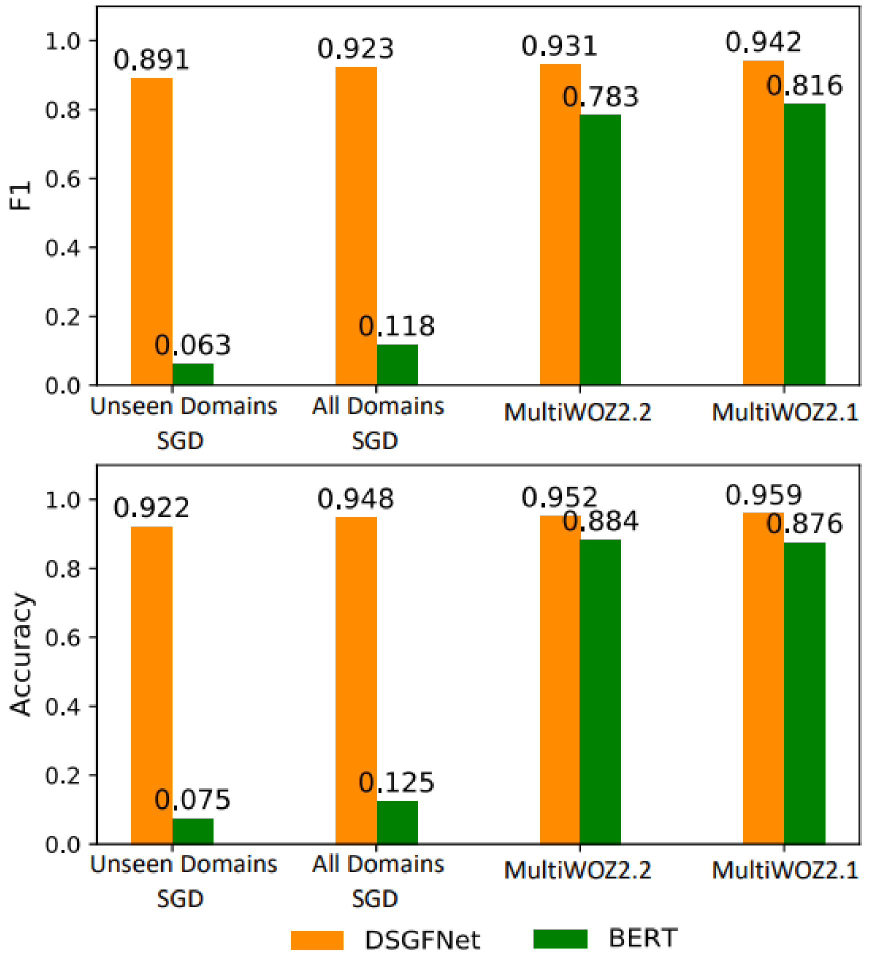


Figure 3: F1 and Accuracy of DSGFNet and BERT for dynamic relation prediction on unseen domains SGD, all domains of SGD, MultiWOZ2.2 and MultiWOZ2.1.

Experiments

Table 6: Case study of DSGFNet and Seq2Seq-DU on SGD. Slot values are extracted from the dialogue context with the same color. The relation of yellow high-light slot pair is predicted as co-reference. The relation of red underline slot pair is predicted as co-update. The relation of bold font slot pair is predicted as co-occurrence. Slot values in red high-light are incorrectly predicted ones.

Dialogue Utterance	[*User*]: What's the weather going to be like in vancouver on March 10th? [*Sys*]: The forecast average is 68 degrees with a 25 per cent chance of rain. [*User*]: Any good attractions in town? [*Sys*]: I have 10 good options including Bloedel Conservatory, a city park. [*User*]: Lovely! Can you book me a ride there? [*Sys*]: Do you want a luxury or pool ride? How many people? [*User*]: Just a regular ride please, book for 1. [*Sys*]: Confirming you want to book a regular cab to Bloedel Conservatory for 1 person.
Ground Truth Dialogue State	[*Weather*]: city = "vancouver"; date = "March 10th"; [*Travel*]: location = "vancouver"; [*RideSharing*]: destination = "Bloedel Conservatory"; number of seats = "1"; ride type = "regular";
State Predictions of DSGFNet	[*Weather*]: **city** = "vancouver"; **date** = "March 10th"; [*Travel*]: location = "vancouver"; [*RideSharing*]: destination = "Bloedel Conservatory"; number of seats = "1"; ride type = "regular";
State Predictions of Seq2seq-DU	[*Weather*]: city = "vancouver"; date = "March 10th"; [*Travel*]: location= "town"; [*RideSharing*]: destination = "Bloedel Conservatory"; number of seats = "1"; ride type = none;

Experiments

Table 7: Performance comparison with different dynamic slot relations and fully-connected relations on unseen domains of SGD, all domains of SGD, MultiWOZ2.2 and MultiWOZ2.1.

Model	Joint GA Unseen Domains SGD	Joint GA All Domains SGD	Joint GA MultiWOZ 2.2	Joint GA MultiWOZ 2.1
-w All Dynamic Relations	24.4%	32.1%	55.8%	56.7%
-w Co-reference Relation	21.5%	29.8%	53.9%	54.7%
-w Co-occurrence Relation	23.8%	31.7%	55.3%	55.9%
-w Co-update Relation	22.3%	30.1%	53.5%	54.5%
-w/o Dynamic Relations	20.6%	28.6%	52.2%	53.2%
-w Fully-connected Relations	21.3%	29.9%	54.2%	54.9%

Table 8: The proportion of different types of dynamic slot relations on SGD, MultiWOZ2.2, and MultiWOZ2.1 in training sets.

Relation	SGD	MultiWOZ2.2	MultiWOZ2.1
Co-reference	5.11%	4.21%	4.29%
Co-update	9.31%	4.01%	4.13%
Co-occurrence	31.13%	37.53%	36.53%

Experiments

Table 9: Accuracy of DSGFNet in each domain on SGD test set. Domains marked with '*' are those for which the schemata in the test set are not present in the training set. Domains marked with '**' have both the unseen and seen schemata. For other domains, the schemata in the test set are also seen in the training set.

Domain	Joint GA	Domain	Joint GA
*RentalCars**	5.11%	*Homes*	22.46%
*Messaging**	5.48%	*Events**	32.02%
*Payment**	7.31%	*Hotels***	33.13%
*Music**	11.87%	*Movies***	42.13%
*Buses**	12.72%	*Services***	45.39%
*Trains**	16.39%	*Travel*	48.30%
*Flights**	16.64%	*Alarm**	53.27%
*Restaurants**	17.01%	*RideSharing*	56.42%
*Media**	20.83%	*Weather*	68.49%

Experiments

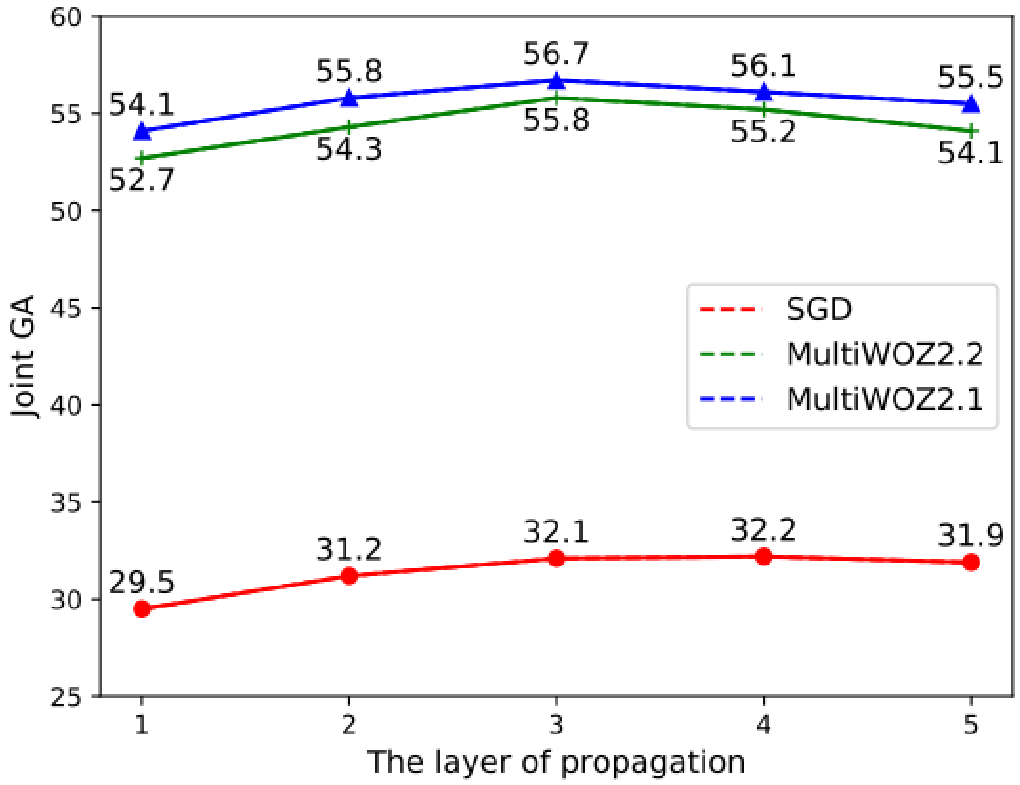


Figure 4: Performance comparison *w.r.t.* the layer of propagation on the schema graph.

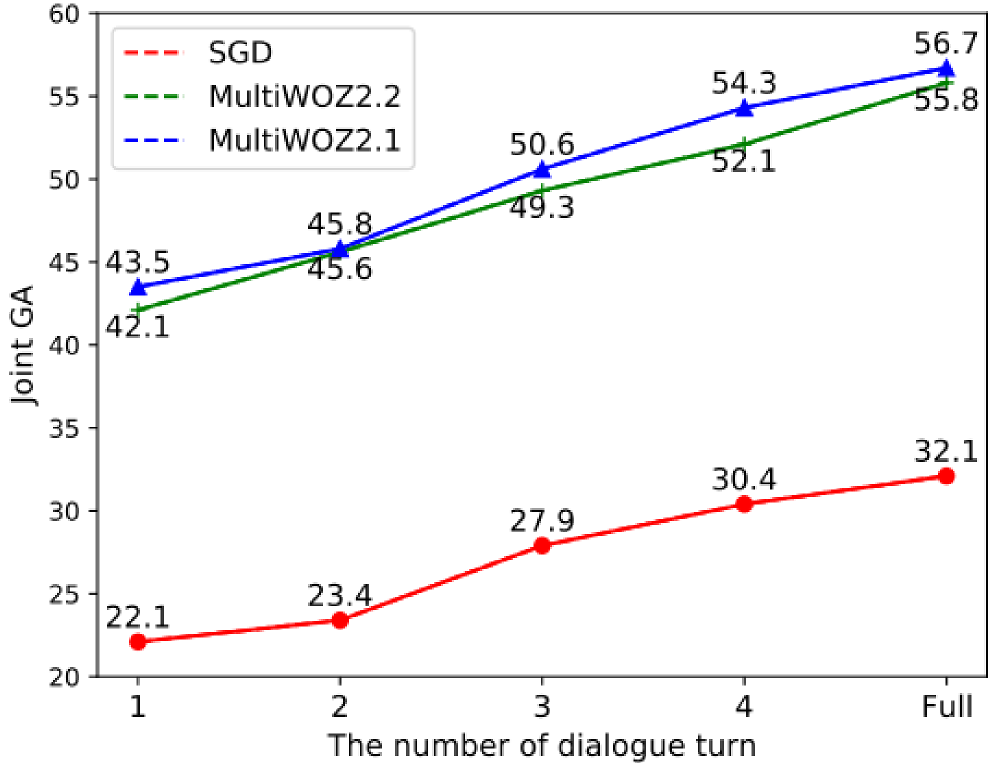


Figure 5: Performance comparison *w.r.t.* the number of dialogue turns used in the schema-dialogue fusion layer.

Experiments

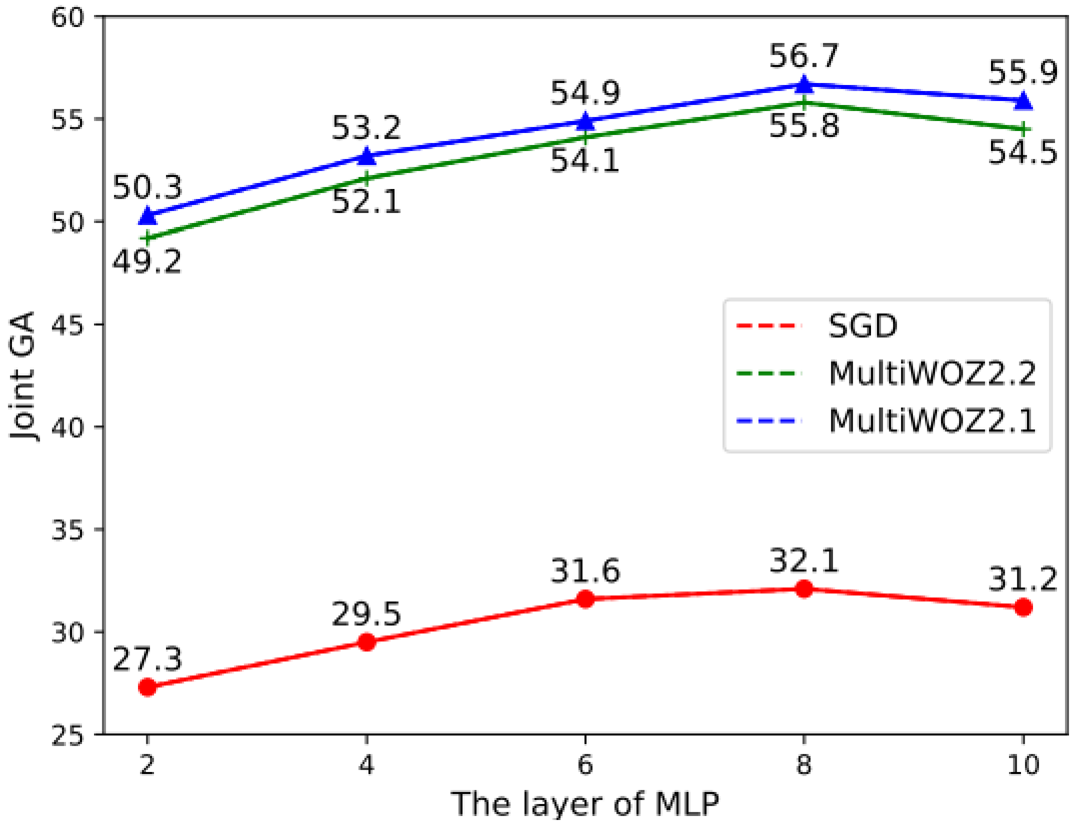


Figure 6: Performance comparison *w.r.t.* the layer of MLP in the dynamic slot relation completion layer.

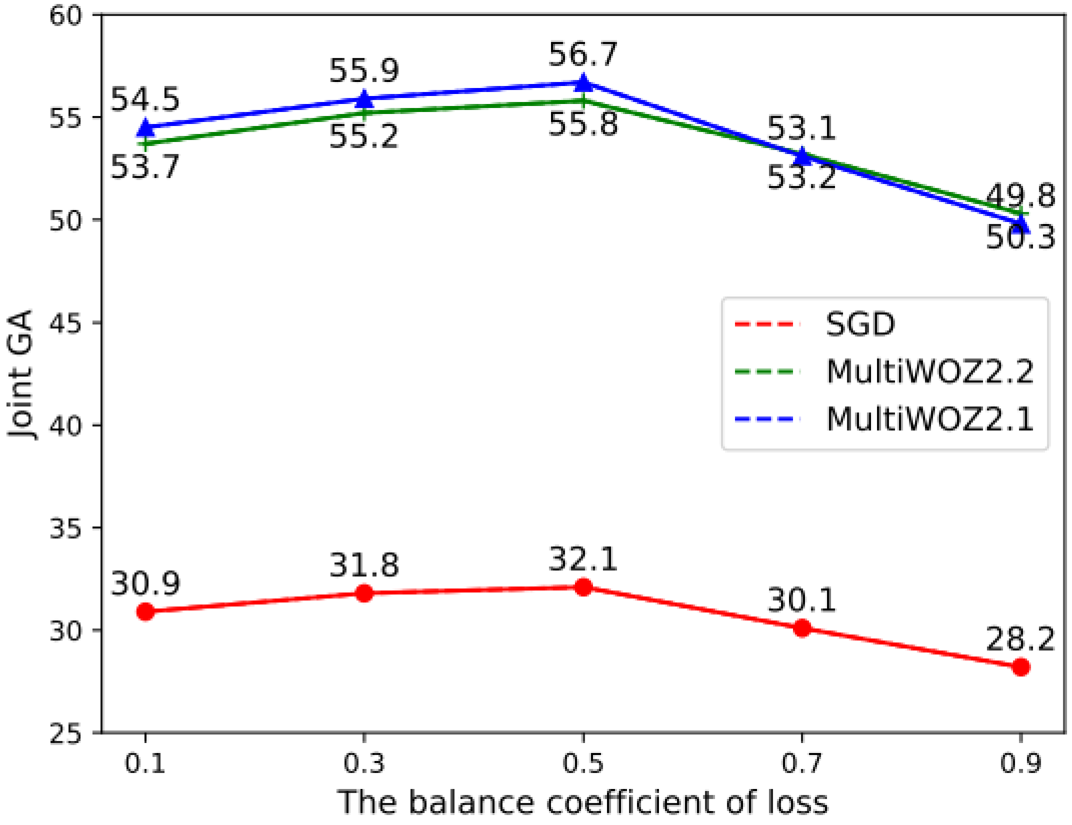


Figure 7: Performance comparison *w.r.t.* the balance coefficient in the loss function.

Thank you！