Co-guiding Net: Achieving Mutual Guidance between Multiple Intent Detection and Slot Filling via Heterogeneous Semantics-Label Graphs

Bowen Xing[1,2] and **Ivor W. Tsang**[2,1]
[1] Australian Artificial Intelligence Institute, University of Technology Sydney, Australia
[2] Centre for Frontier Artificial Intelligence Research, A*STAR, Singapore
bwxing714@gmail.com, ivor.tsang@gmail.com

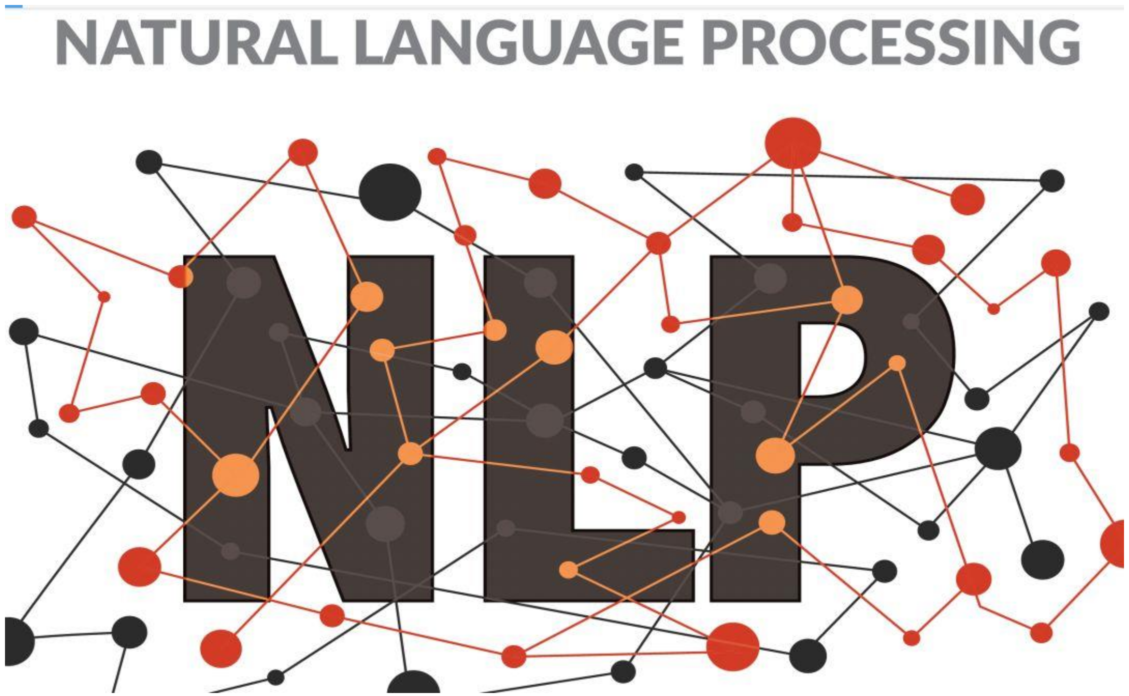
2022.10.24 • ChongQing

2022_EMNLP

Reported by Yidan Liu

gesis Leibniz-Institut für Sozialwissenschaften

Forschungszentrum · Research Center L3S



1.Introduction

2.Method

3.Experiments

Introduction

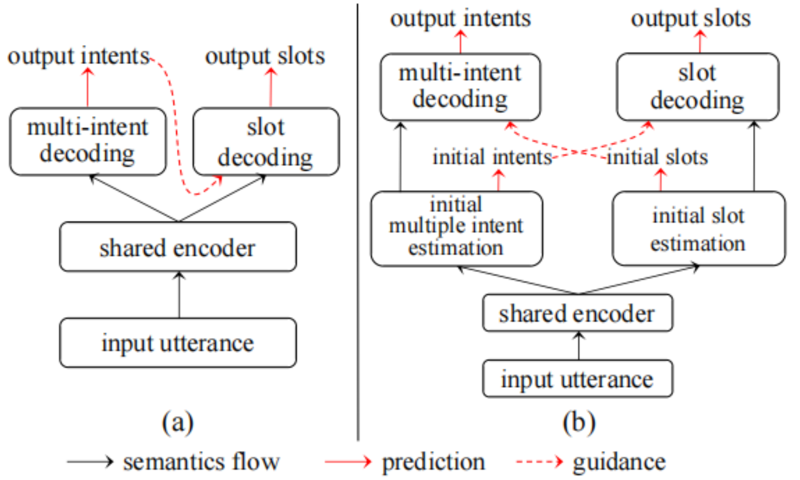
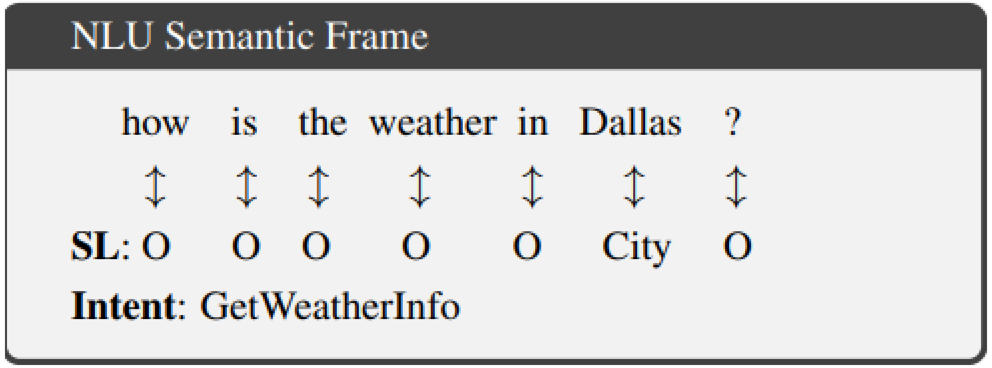


Figure 1: (a) Previous framework which only models the unidirectional guidance from multi-intent predictions to slot filling. (b) Our framework which models the mutual guidances between the two tasks.



- only model the unidirectional guidance from intent to slot
- adopt homogeneous graphs to model the interactions between the slot semantics nodes and intent label nodes

Method

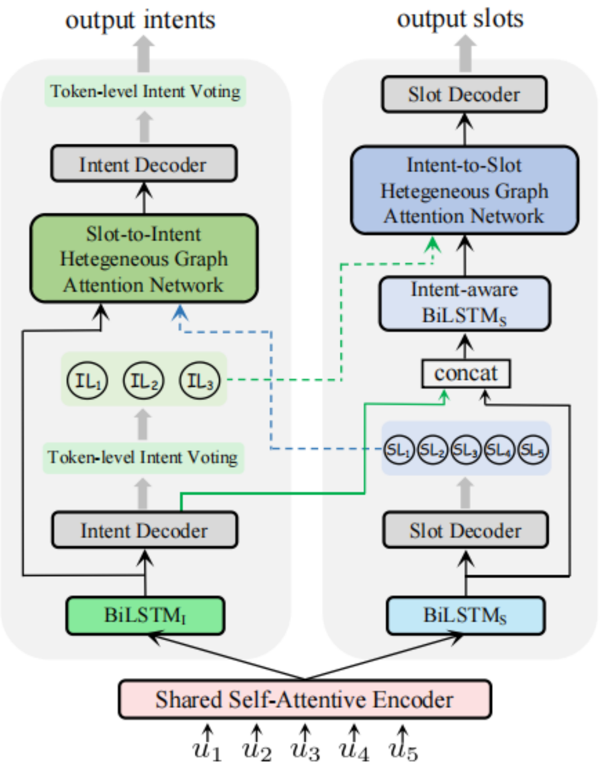


Figure 5: The architecture of Co-guiding Net. Each HGAT is triggered by its own task's semantics and the counterpart's predict labels. The green and blue dashed arrow lines denote the projected label representations from the predicted intents and slots, respectively. The green solid arrow line denotes the intent distribution generated by the Intent Decoder at the first stage.

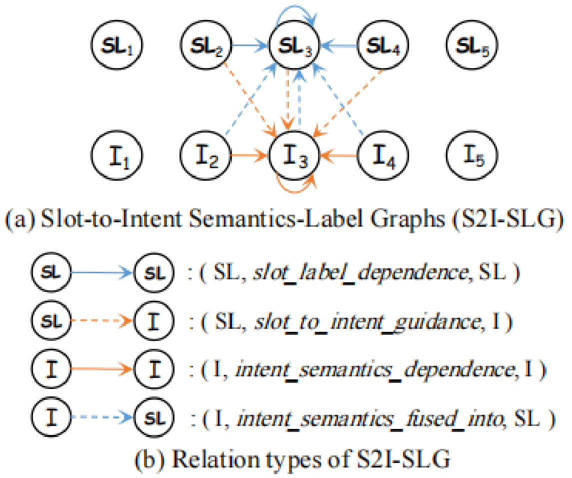


Figure 3: The illustration of S2I-SLG and its relation types. w.l.o.g, only the edges directed into SL_3 and I_3 are shown, and the local window size is 1.

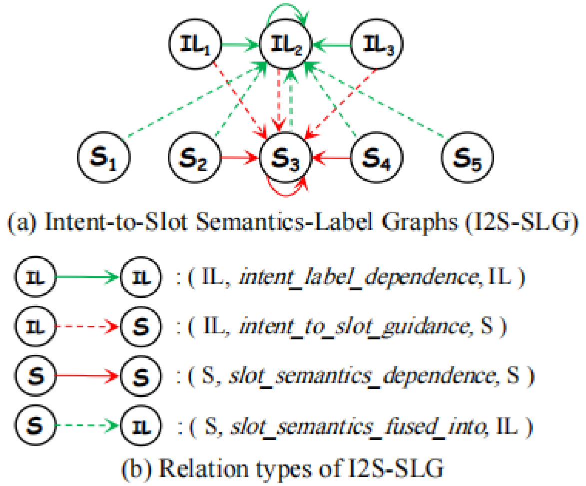
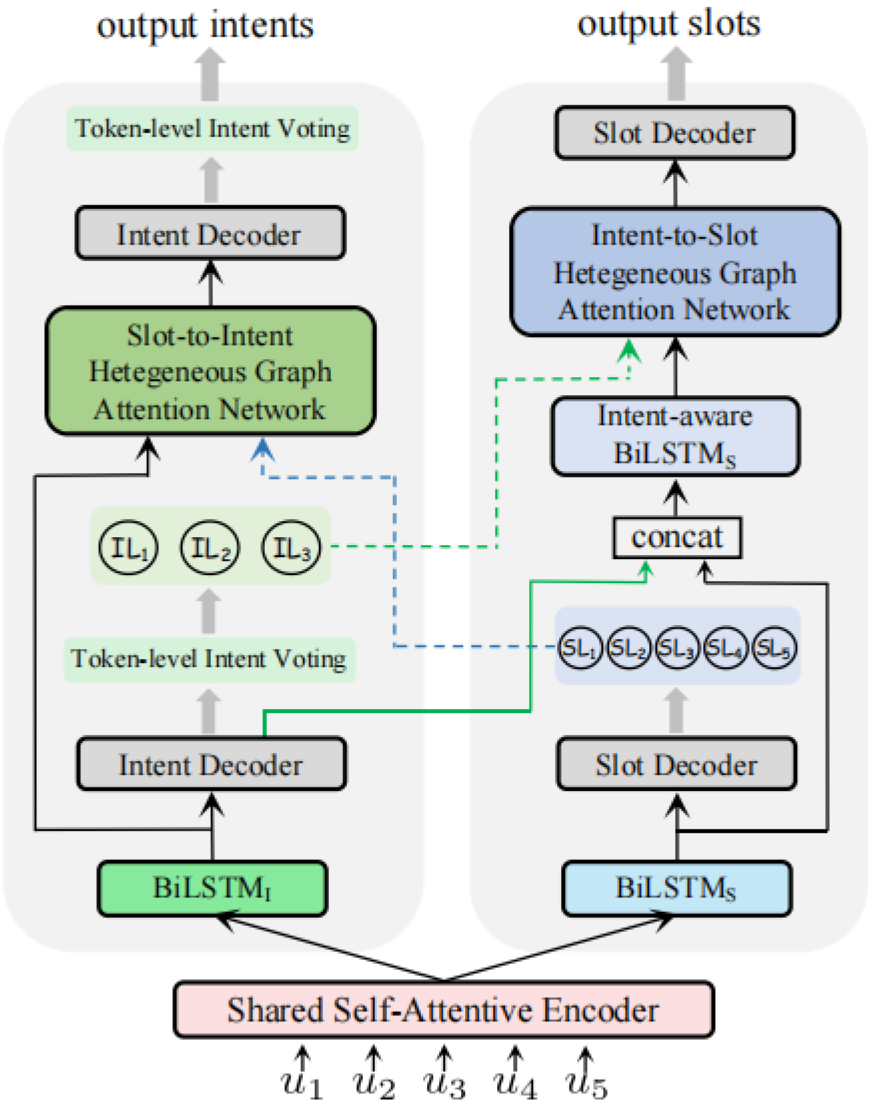


Figure 4: The illustration of I2G-SLG and its relation types. w.l.o.g, only the edges directed into IL_3 and S_3 are shown, and the local window size is 1.

Method



Shared Self-Attentive Encoder

$$h_i = \mathrm{BiLSTM}\left(x_i, h_{i-1}, h_{i+1}\right) \quad (1)$$

$$\boldsymbol{H}' = \mathrm{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^\top}{\sqrt{d_k}}\right)\boldsymbol{V} \quad (2)$$

$$\boldsymbol{H} = \hat{\boldsymbol{H}} \| \boldsymbol{H}'$$

Initial Estimation

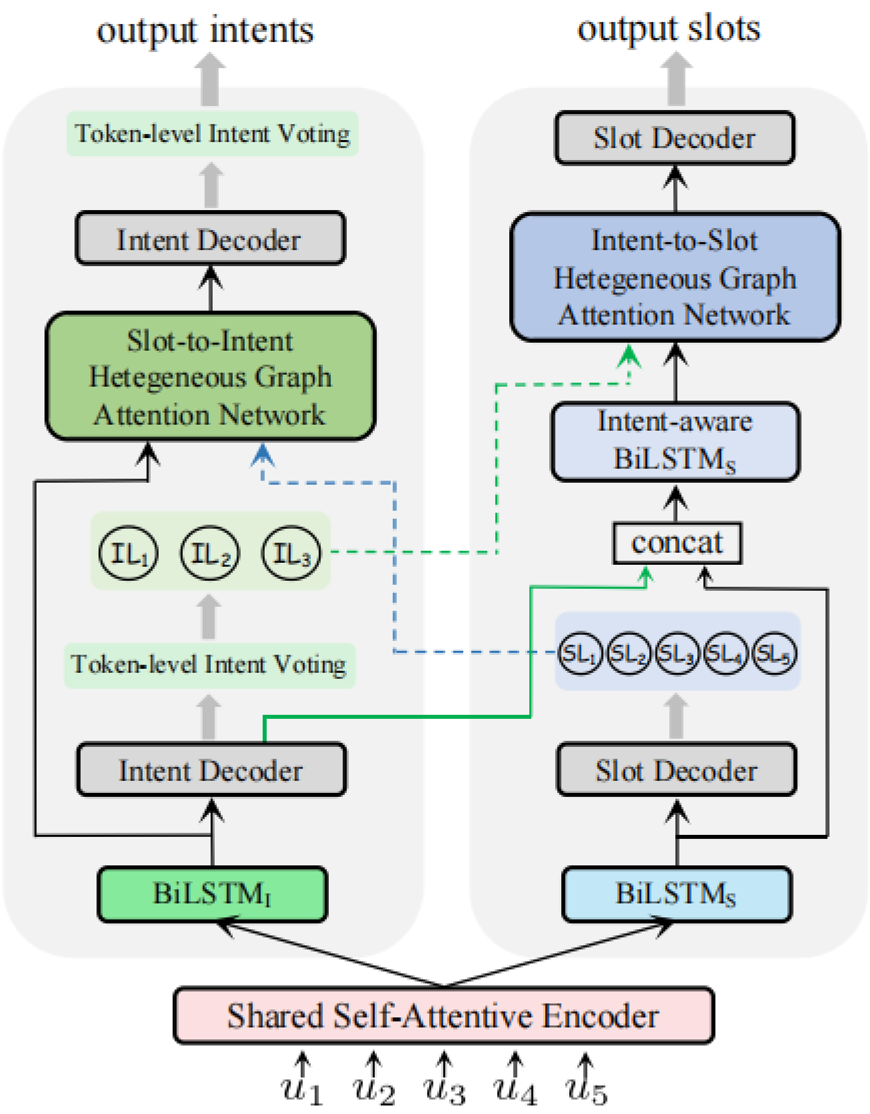
$$h_i^{[I,0]} = \mathrm{BiLSTM_I}\left(h_i, h_{i-1}^{[I,0]}, h_{i+1}^{[I,0]}\right) \quad (3)$$

$$y_i^{[I,0]} = \mathrm{sigmoid}\left(\boldsymbol{W}_I^1\left(\sigma(\boldsymbol{W}_I^2 \boldsymbol{h}_i^{[I,0]} + \boldsymbol{b}_I^2)\right) + \boldsymbol{b}_I^1\right) \quad (4)$$

$$h_i^{[S,0]} = \mathrm{BiLSTM_S}(h_i, h_{i-1}^{[S,0]}, h_{i+1}^{[S,0]}) \quad (5)$$

$$y_i^{[S,0]} = \mathrm{softmax}\left(\boldsymbol{W}_S^1\left(\sigma(\boldsymbol{W}_S^2 \boldsymbol{h}_i^{[S,0]} + \boldsymbol{b}_S^2)\right) + \boldsymbol{b}_S^1\right) \quad (6)$$

Method



Heterogeneous Graph Attention Network

$$h_i^{l+1} = \Big\|_{k=1}^{K} \sigma\left(\sum_{j\in\mathcal{N}_{s2i}^{i}} W_{s2i}^{[r,k,1]} \alpha_{ij}^{[r,k]} h_j^l\right), r = \phi\left(e_{s2i}^{[j,i]}\right)$$

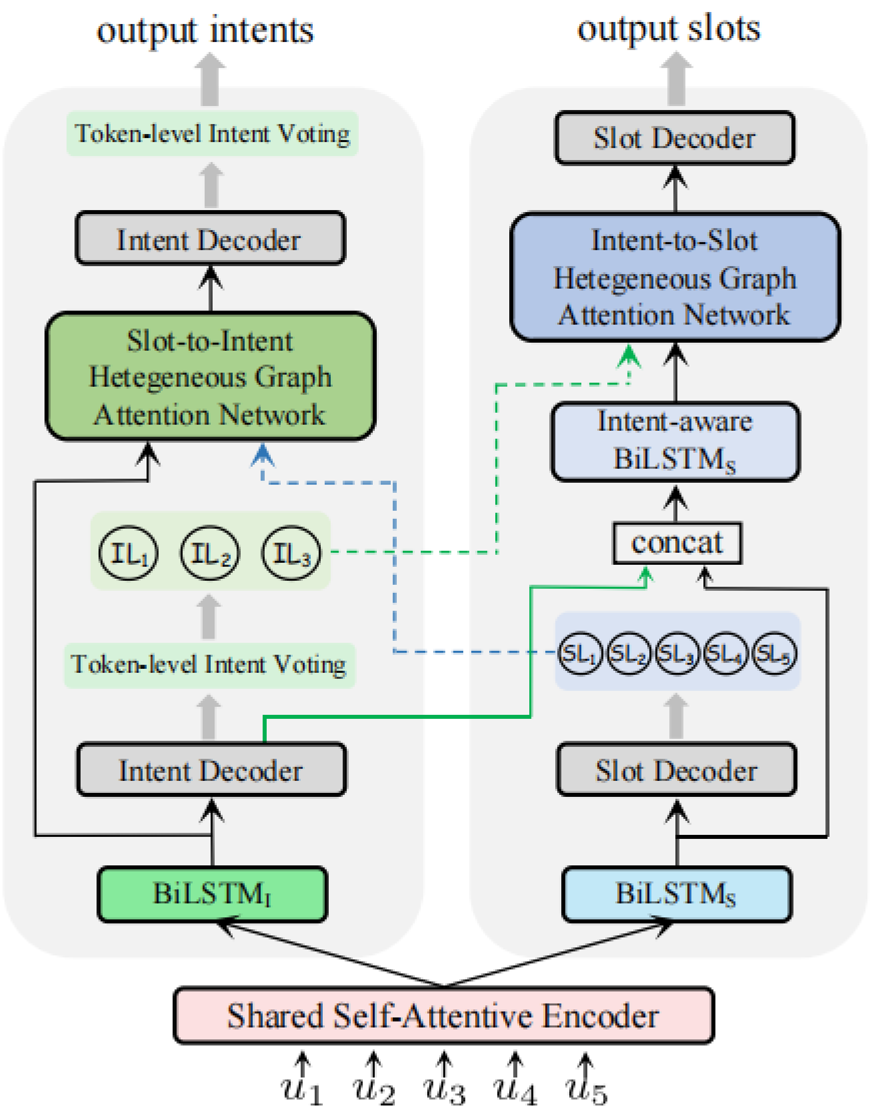
$$\alpha_{ij}^{[r,k]} = \frac{\exp\left(\left(W_{s2i}^{[r,k,2]} h_i^l\right)\left(W_{s2i}^{[r,k,3]} h_j^l\right)^{\mathsf{T}} / \sqrt{d}\right)}{\sum_{u\in\mathcal{N}_{s2i}^{r,i}} \exp\left(\left(W_{s2i}^{[r,k,2]} h_i^l\right)\left(W_{s2i}^{[r,k,3]} h_u^l\right)^{\mathsf{T}} / \sqrt{d}\right)} \qquad (7)$$

$$H^{[I,L]} = \text{S2I-HGAT}\left([H^{[I,0]}, E_{sl}], \mathcal{G}_{s2i}, \theta_I\right) \qquad (8)$$

$$\tilde{h}_i^{[S,0]} = \text{BiLSTM}(y_i^{[I,0]} \| h_i^{[S,0]}, \tilde{h}_{i-1}^{[S,0]}, \tilde{h}_{i+1}^{[S,0]}) \qquad (9)$$

$$H^{[S,L]} = \text{I2S-HGAT}\left([\tilde{H}^S, E_{il}], \mathcal{G}_{i2s}, \theta_S\right) \qquad (10)$$

Method



Loss Function

$$\mathrm{CE}(\hat{y}, y) = \hat{y}\log(y) + (1-\hat{y})\log(1-y)$$
$$\mathcal{L}_I = \sum_{t=0}^{1}\sum_{i=1}^{n}\sum_{j=1}^{N_I} \mathrm{CE}\left(\hat{y}_i^I[j], y_i^{[I,t]}[j]\right) \tag{11}$$

$$\mathcal{L}_S = \sum_{t=0}^{1}\sum_{i=1}^{n}\sum_{j=1}^{N_S} \hat{y}_i^S[j]\log\left(y_i^{[S,t]}[j]\right) \tag{12}$$

Margin Penalty

$$\mathcal{L}_I^{mp} = \sum_{i=1}^{n}\sum_{j=1}^{N_I} \hat{y}_i^I[j]\ \max\left(0, y_i^{[I,0]}[j] - y_i^{[I,1]}[j]\right)$$
$$\mathcal{L}_S^{mp} = \sum_{i=1}^{n}\sum_{j=1}^{N_S} \hat{y}_i^S[j]\max\left(0, y_i^{[S,0]}[j] - y_i^{[S,1]}[j]\right) \tag{13}$$

$$\mathcal{L} = \gamma\left(\mathcal{L}_I + \beta_I\mathcal{L}_I^{mp}\right) + (1-\gamma)\left(\mathcal{L}_S + \beta_S\mathcal{L}_S^{mp}\right) \tag{14}$$

Experiment

Models	MixATIS			MixSNIPS		
	Overall(Acc)	Slot (F1)	Intent(Acc)	Overall(Acc)	Slot(F1)	Intent(Acc)
Attention BiRNN (Liu and Lane, 2016)	39.1	86.4	74.6	59.5	89.4	95.4
Slot-Gated (Goo et al., 2018)	35.5	87.7	63.9	55.4	87.9	94.6
Bi-Model (Wang et al., 2018)	34.4	83.9	70.3	63.4	90.7	95.6
SF-ID (E et al., 2019)	34.9	87.4	66.2	59.9	90.6	95.0
Stack-Propagation (Qin et al., 2019)	40.1	87.8	72.1	72.9	94.2	96.0
Joint Multiple ID-SF (Gangadharaiah and Narayanaswamy, 2019)	36.1	84.6	73.4	62.9	90.6	95.1
AGIF (Qin et al., 2020)	40.8	86.7	74.4	74.2	94.2	95.1
GL-GIN (Qin et al., 2021b)	43.0	88.2	76.3	73.7	94.0	95.7
Co-guiding Net (ours)	**51.3**†	**89.8**†	**79.1**†	**77.5**†	**95.1**†	**97.7**†

Table 1: Results comparison. † denotes our model significantly outperforms baselines with $p < 0.01$ under t-test.

Experiment

Models	MixATIS			MixSNIPS		
	Overall(Acc)	Slot (F1)	Intent(Acc)	Overall(Acc)	Slot(F1)	Intent(Acc)
Co-guiding Net	**51.3**	**89.8**	**79.1**	**77.5**	**95.1**	**97.7**
w/o S2I-guidance	47.7 (↓3.6)	88.8 (↓1.0)	77.1 (↓2.0)	76.6 (↓0.9)	94.7 (↓0.4)	96.9 (↓0.8)
w/o I2S-guidance	47.7 (↓3.6)	88.7 (↓1.1)	77.5 (↓1.6)	76.5 (↓1.0)	94.9 (↓0.2)	97.5 (↓0.2)
w/o relations	46.0 (↓5.3)	88.3 (↓1.5)	77.8 (↓1.3)	76.3 (↓1.2)	94.7 (↓0.5)	97.2 (↓0.4)
+ Local Slot-aware GAT	51.1 (↓0.2)	89.4 (↓0.4)	79.0 (↓0.1)	75.9 (↓1.6)	94.7 (↓0.4)	96.4 (↓1.4)

Table 2: Results of ablation experiments.

Experiment

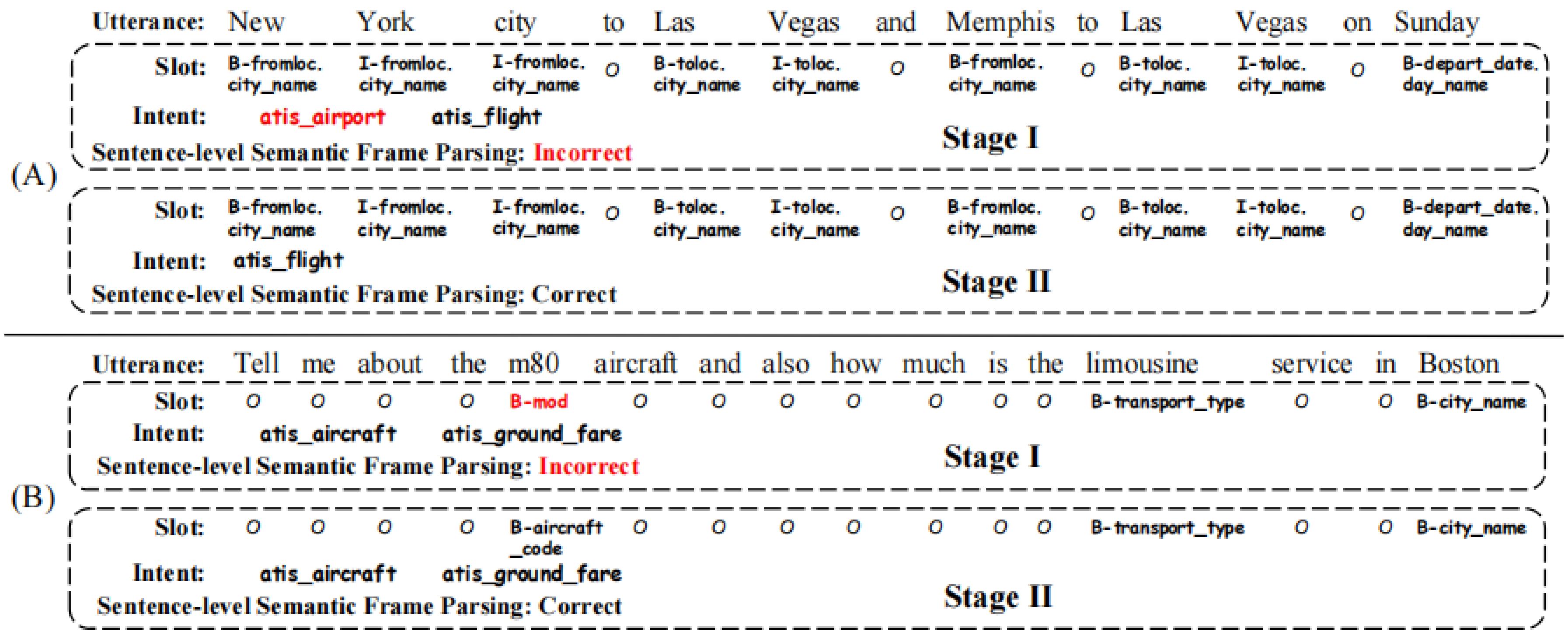


Figure 6: Case study of slot-to-intent guidance (A) and intent-to-slot guidance (B). Red color denotes error.

Thank you!