Dynamic Time-Aware Attention to Speaker Roles and Contexts for Spoken Language Understanding

miu

Po-Chun Chen, Ta-Chung Chi*, Shang-Yu Su*, Yun-Nung (Vivian) Chen*

*Three authors contribute this work equally.

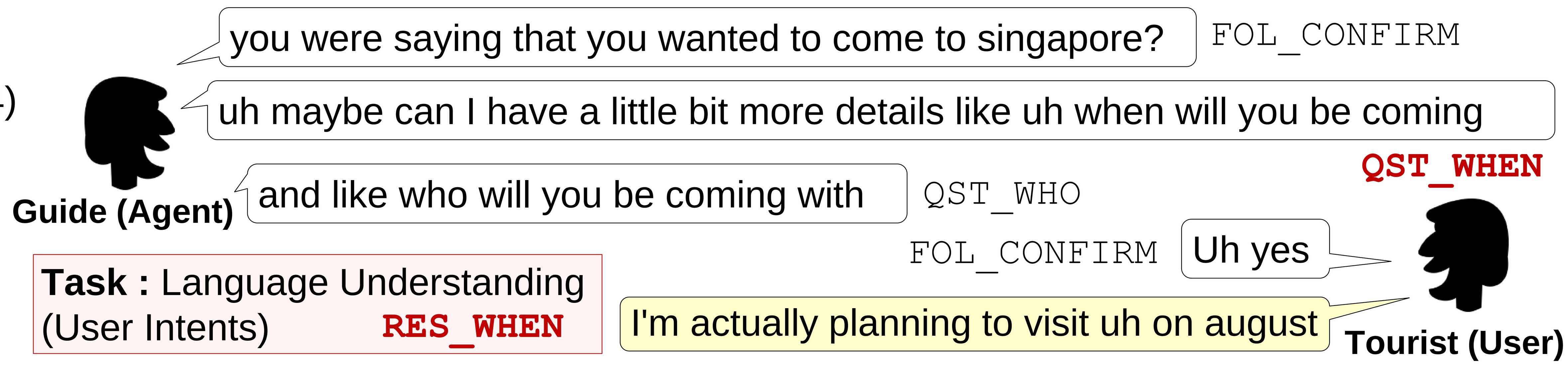
國立臺灣大學 National Taiwan University

https://github.com/MiuLab/Time-SLU

Summary

➢Task Definition
- Language understanding for human-human dialogues between tourists and guides (DSTC4)

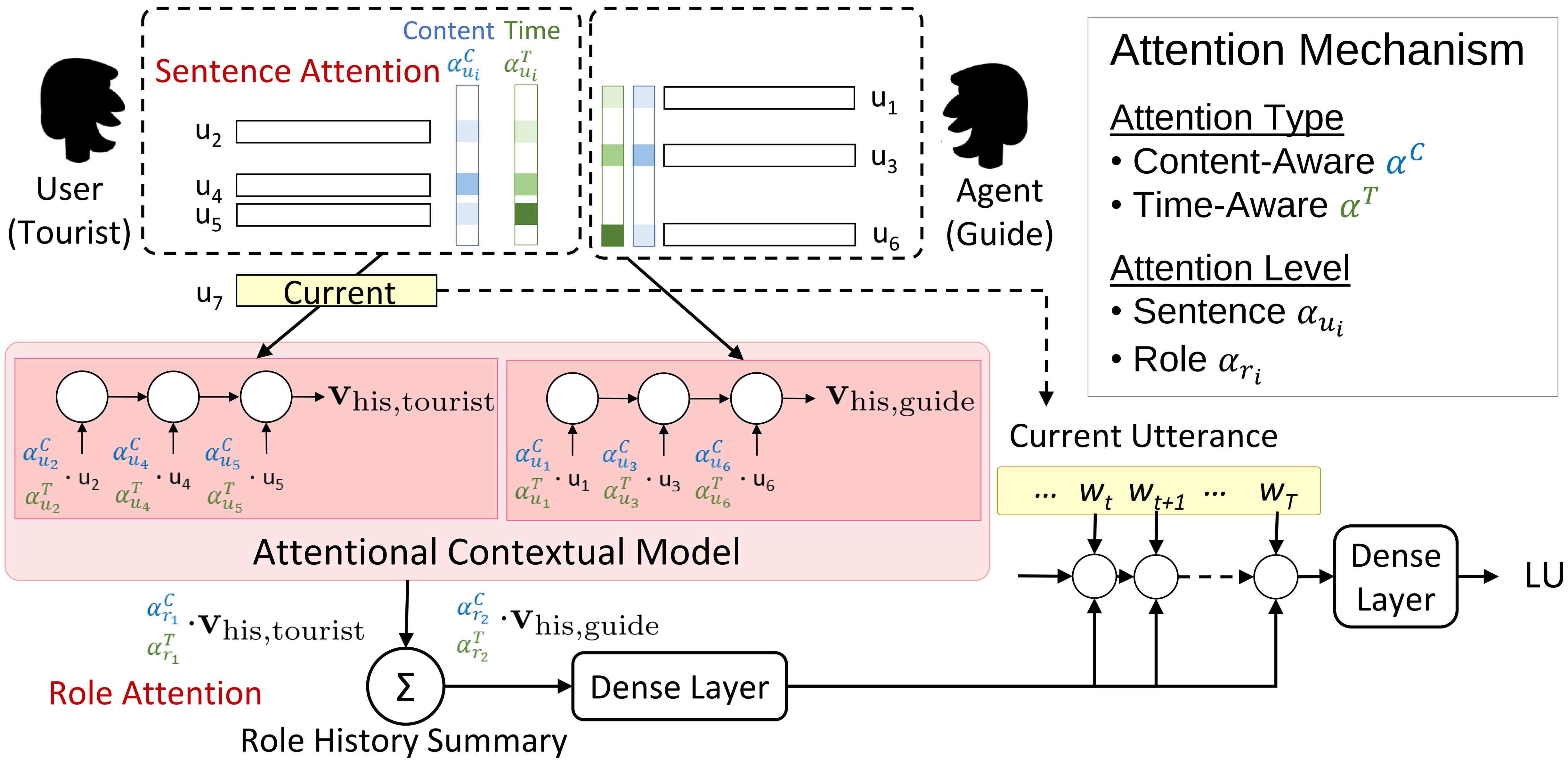
➢Motivation
- Human-human dialogues contain multiple reasoning steps
- Additional attention mechanism like ***role*** or ***temporal*** information may be useful



➢Method: ***Time-Aware Attention to Speaker Roles and Contexts***
- Modeling by the fixed time-aware attention and other learnable attention mechanisms

➢Result
- The model achieves state-of-the-art performance

The Proposed Approach: Time-Aware Attention to Roles & Contexts



Attention Mechanism

Attention Type
- Content-Aware α^C
- Time-Aware α^T

Attention Level
- Sentence α_{u_i}
- Role α_{r_i}

➢***Time-Aware Attention***
- Recent utterance contains more relevant information

$$\alpha_u^T = \frac{1}{d(u)}$$

distance between *u* and the current utterance

➢***Content-Aware Attention***
- The semantic relation decides where and how much the model should focus on given the contexts

$$\alpha_u^C = \mathtt{softmax}(M_1(\mathbf{v}_{\text{cur}} + \mathbf{v}_u))$$

➢***Role-Level Attention***
- Different speaker roles behave differently
- Role-level attention is based on how much to address on different speaker roles' contexts

$$\alpha_r = \mathtt{softmax}(M_2(\mathbf{v}_{\text{cur}} + \mathbf{v}_{\text{his,r}}))$$

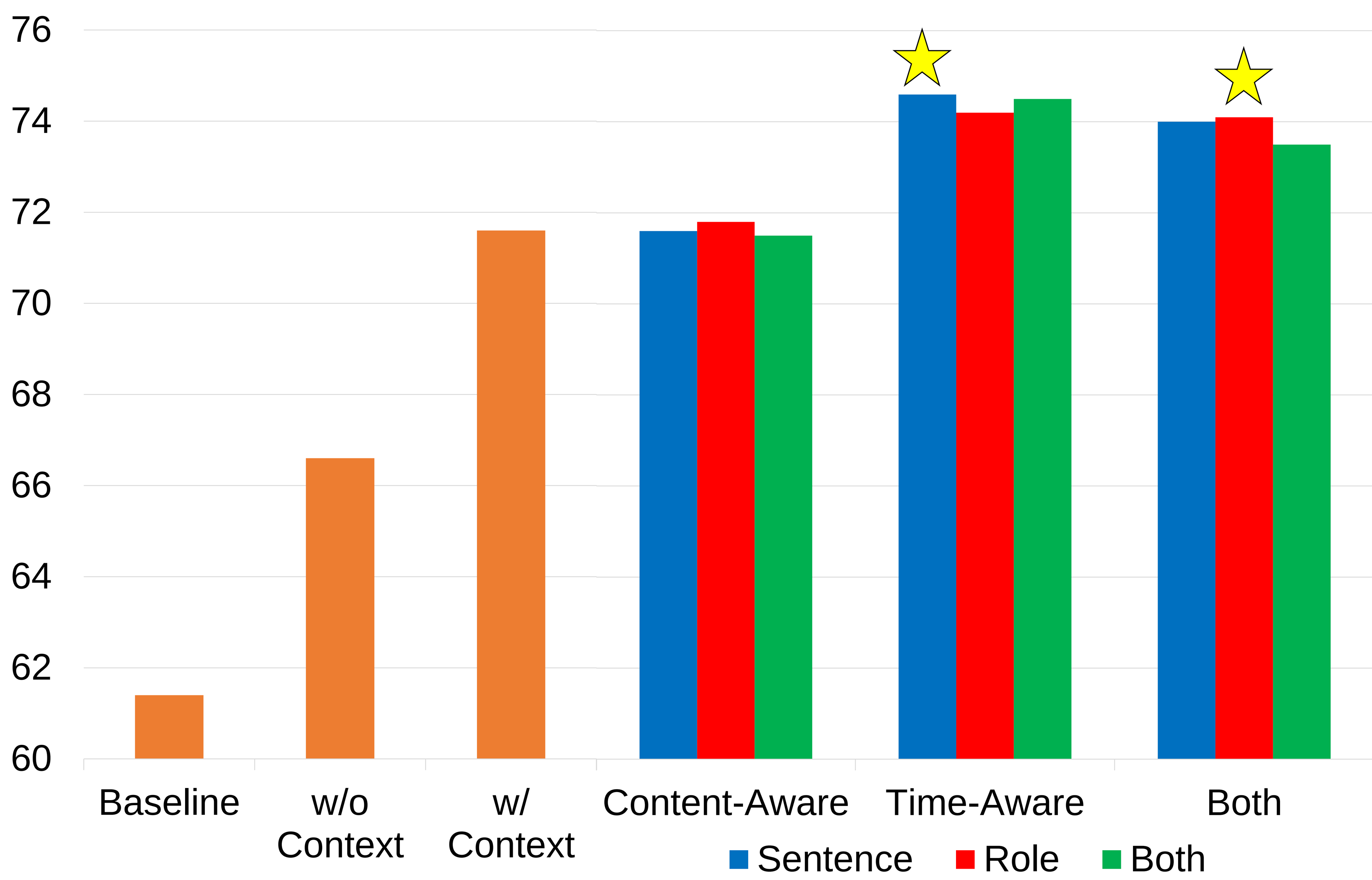
➢***End-to-End Training for Contextual Language Understanding***
- BLSTM-encoded current utterance concatenated with the history vector for *multi-label intent prediction*

$$\vec{y} = \text{BLSTM}(\vec{v}_{\text{his}}, \vec{x}) \quad p(\vec{y} \mid \vec{x}) = \prod_i p(y_i \mid w_1, \cdots, w_i)$$

- All encoders, prediction models, and attention weights (except time-aware attention) can be automatically learned in an end-to-end manner $\mathbf{v}_{\text{his}} = \sum_i \alpha_{\text{role}_i}^{C/T} \cdot \mathbf{v}_{\text{his,role}_i}$

➢ Leveraging different types and levels of attention can improve language understanding

Experiments and Discussions



➢Setup
- Dataset: DSTC4 35 human-human dialogues
- Evaluation metrics: F1 for multi-label classification

➢Experimental Results
- ***Time-aware attention*** models significantly outperform the baselines
- The ***attention mechanisms*** provide improvement for LU
- Role-level attention requires content-related information to achieve better performance

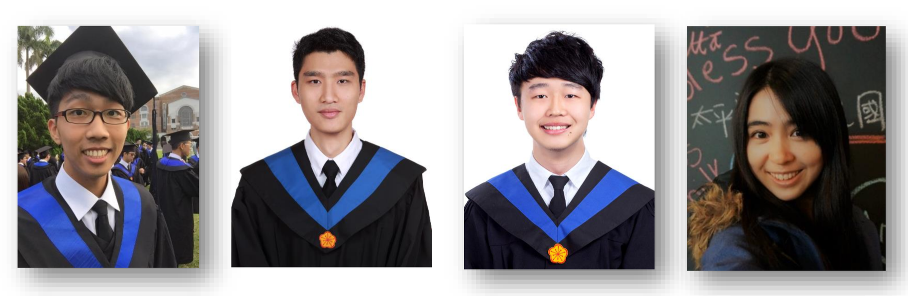
➢Discussion
- Guide results are consistently better than tourist results
- The reason may be that the guide has similar behavior patterns (e.g. providing information and confirming questions) while the user has more diverse intentions

➢ Time-aware attention brings significant improvement for language understanding

Conclusions

- **Approach**: an end-to-end attentional role-based contextual model that leverages various content-aware and time-aware attention mechanisms
- **Experiment**: impressive improvement on a benchmark multi-domain dialogue dataset
- **Result**: temporal information is important when modeling history utterances for language understanding

b02902019@ntu.edu.tw
r06922028@ntu.edu.tw
r05921117@ntu.edu.tw
y.v.chen@ieee.org



Code Available:
https://github.com/MiuLab/Time-SLU