Speaker Role Contextual Modeling for Language Understanding and Dialogue Policy Learning

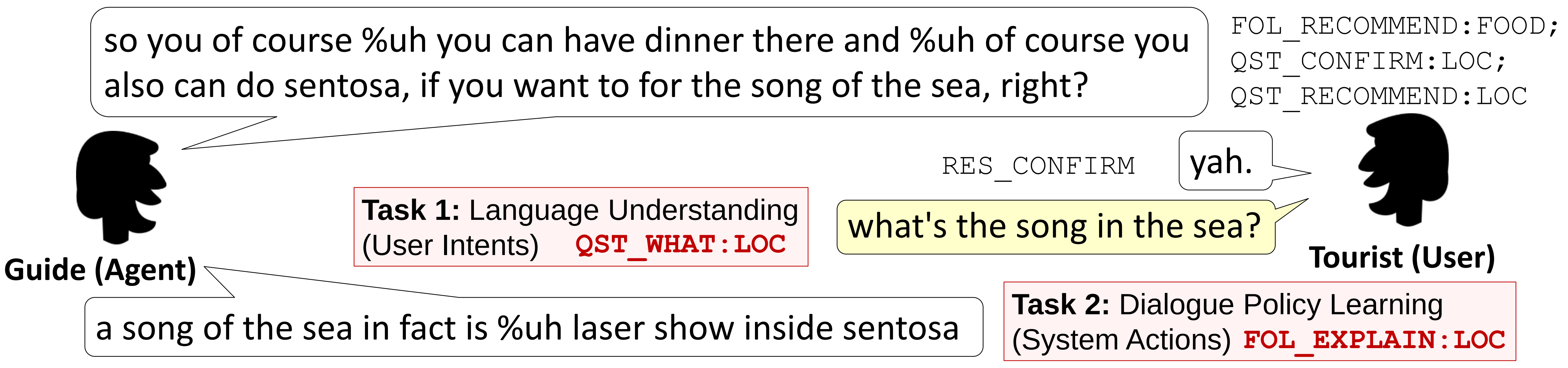
National Taiwan University

Ta-Chung Chi, Po-Chun Chen*, Shang-Yu Su*, Yun-Nung (Vivian) Chen* **Three authors contribute this work equally.*

https://github.com/MiuLab/Spk-Dialogue

Summary

- Task Definition
  - DSTC4: human-human dialogues between tourists and guides
- Motivation
  - Human-human dialogues contain ***rich and complex*** human behaviors
  - ***Different speaker roles*** behave differently and cause notable variance in speaking habits
- Method: ***Role-Based Contextual Model for LU & PL***
  - Introduce two separate models to represent two speaker roles
- Result
  - The model achieves impressive improvement on the DSTC4 dataset



so you of course %uh you can have dinner there and %uh of course you also can do sentosa, if you want to for the song of the sea, right?

```
FOL_RECOMMEND:FOOD;
QST_CONFIRM:LOC;
QST_RECOMMEND:LOC
```

Guide (Agent)

Tourist (User)

RES_CONFIRM yah.

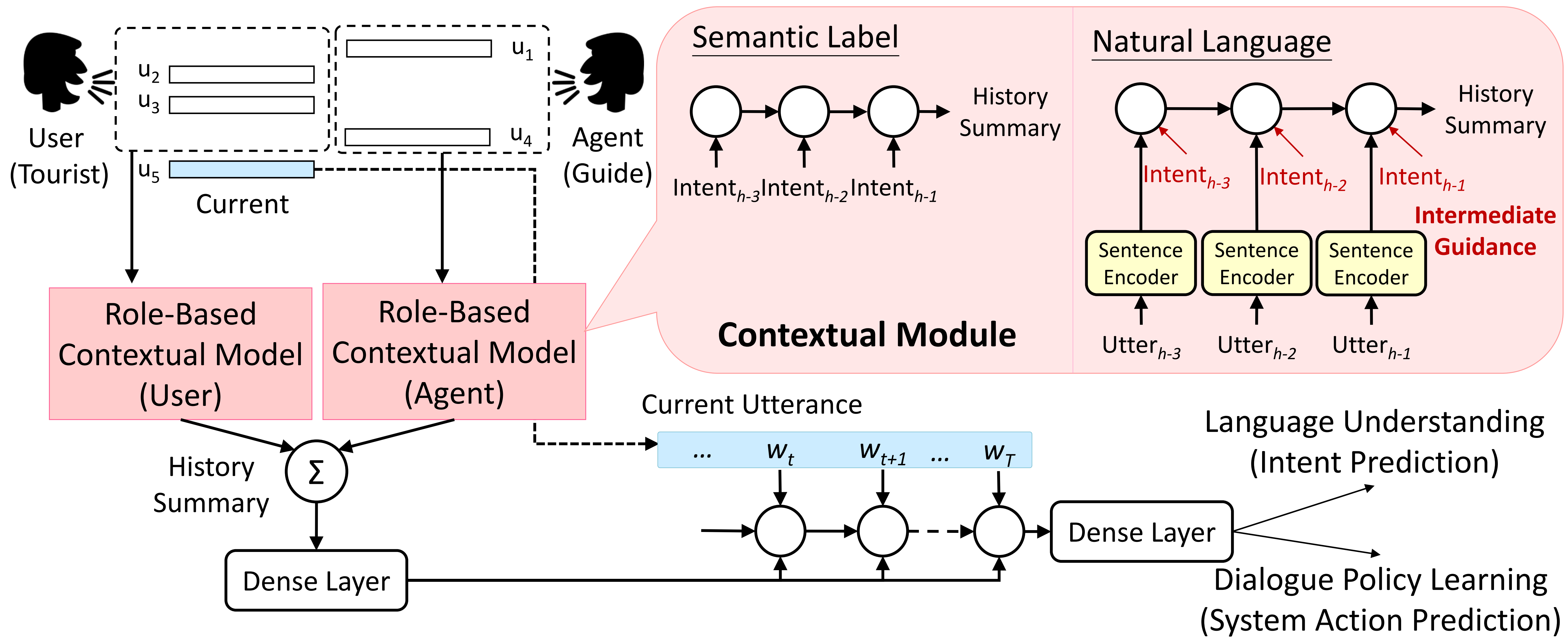
what's the song in the sea?

Task 1: Language Understanding (User Intents) **QST_WHAT:LOC**

a song of the sea in fact is %uh laser show inside sentosa

Task 2: Dialogue Policy Learning (System Actions) **FOL_EXPLAIN:LOC**

The Proposed Approach: Role-Based Model for LU & PL



User (Tourist)
u_2
u_3
u_1
u_4
Agent (Guide)
u_5
Current
Role-Based Contextual Model (User)
Role-Based Contextual Model (Agent)
Semantic Label
History Summary
Intent$_{h-3}$ Intent$_{h-2}$ Intent$_{h-1}$
Natural Language
History Summary
Intent$_{h-3}$ Intent$_{h-2}$ Intent$_{h-1}$
Intermediate Guidance
Sentence Encoder
Sentence Encoder
Sentence Encoder
Utter$_{h-3}$ Utter$_{h-2}$ Utter$_{h-1}$
Contextual Module
Current Utterance
… w_t w_{t+1} … w_T
History Summary
Σ
Dense Layer
Dense Layer
Language Understanding (Intent Prediction)
Dialogue Policy Learning (System Action Prediction)

❖ Contextual Model

- **encoding contexts as a history vector v_{his}**
- ✓ **Semantic Label:** ground-truth intent tags are encoded as the 1-hot sentence semantics

$$\vec{v}_{\text{his}} = \text{BLSTM}(\vec{\text{intent}_t})$$

- ✓ **Natural Language:** CNN-encoded sentence vector for practical situations

$$\vec{v}_{\text{his}} = \text{BLSTM}(\text{CNN}(\vec{\text{utt}_t}))$$

- ✓ **NL w/ Intermediate Guidance:** semantic labels act as middle supervision signal for guiding the sentence encoding module to project from input utterances to a more meaningful feature space

> ➢ Leverage contextual information for better understanding

❖ Speaker Role Modeling

- **train two role-specific models independently, BLSTM$_{\text{rolea}}$ and BLSTM$_{\text{roleb}}$**

$$\vec{v}_{\text{his}} = \text{BLSTM}_{\text{role}_a}(\vec{\text{intent}}_{t,\text{role}_a}) + \text{BLSTM}_{\text{role}_b}(\vec{\text{intent}}_{t,\text{role}_b})$$

$$\vec{v}_{\text{his}} = \text{BLSTM}_{\text{role}_a}(\text{CNN}(\vec{\text{utt}}_{t,\text{role}_a})) + \text{BLSTM}_{\text{role}_b}(\text{CNN}(\vec{\text{utt}}_{t,\text{role}_b}))$$

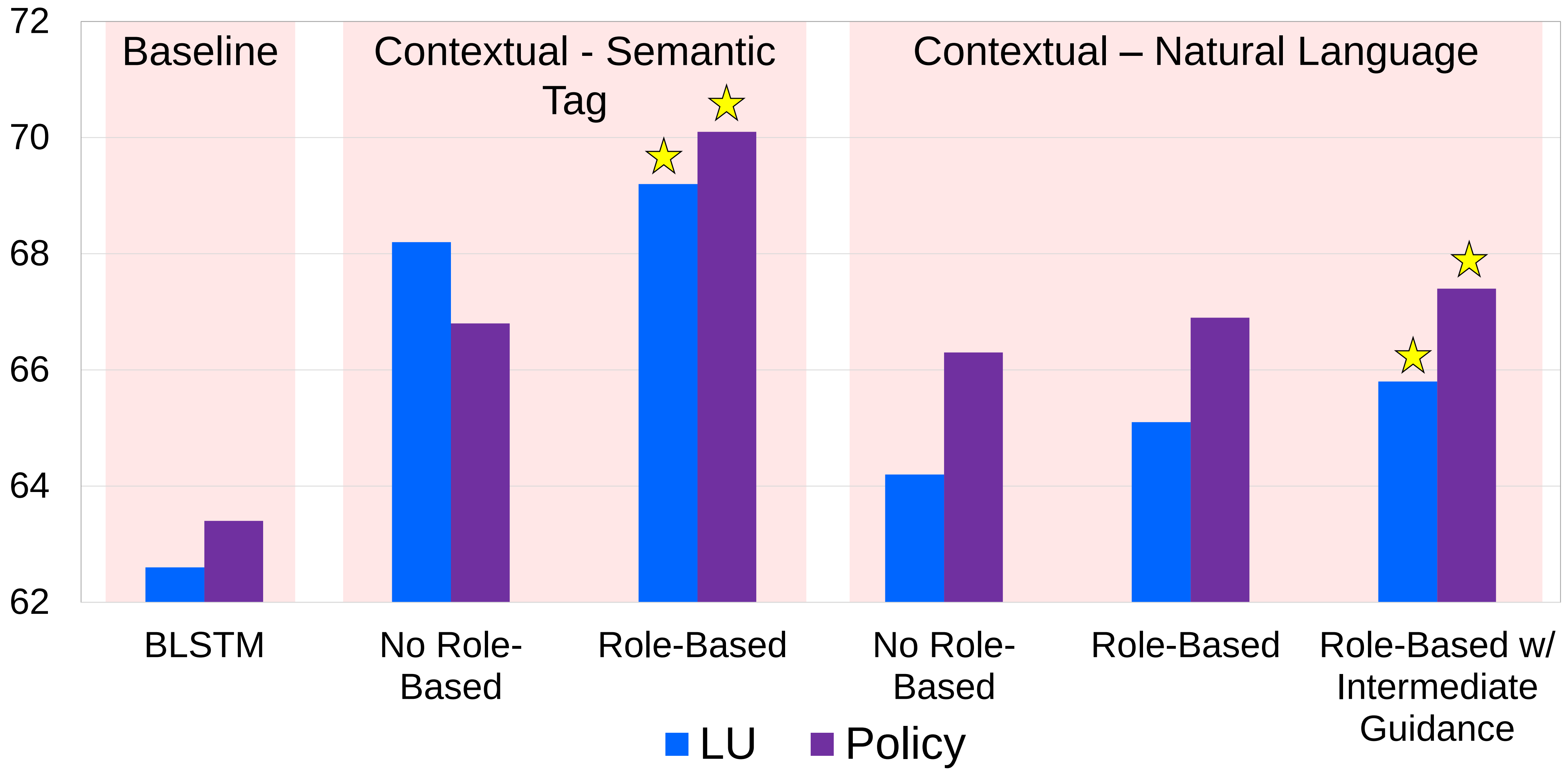
> ➢ User usually pay attention to ***self history (reasoning)*** and ***others' utterances (listening)***
> ➢ Two speaker roles behave differently

- **End-to-End Training Objective**
  - BLSTM-encoded current utterance concatenated with the history vector for *multi-label intent prediction and system action prediction*

$$\vec{y} = \text{BLSTM}(\vec{v}_{\text{his}}, \vec{x})$$

$$p(\vec{y} \mid \vec{x}) = \prod_i p(y_i \mid w_1, \cdots, w_i)$$

Experiments and Discussions



> ➢ The proposed speaker role contextual model obtains the state-of-the-art results.

- Setup
  - Dataset: DSTC4 35 human-human dialogues
  - Evaluation metrics: F1 for multi-label classification
- Experimental Results
  - ***Contextual*** models significantly improve the baselines
  - The ***role-based*** models outperform the one without the role information for both tasks
  - ***Intermediate guidance*** improves semantic modeling
- Discussion
  - Most LU results are worse than dialogue policy learning 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 interactions*
  - The idea about modeling speaker role information can be further extended to various research topics

Conclusions

- **Approach**: an end-to-end role-based contextual model that automatically learns speaker specific contextual encoding
- **Experiment**: impressive improvement on a benchmark multi-domain dialogue dataset
- **Result**: demonstrating that different speaker roles behave differently and focus on different goals

Code Available:
https://github.com/MiuLab/Spk-Dialogue

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