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

The model achieves impressive improvement on the DSTC4 dataset

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

Contextual Model
- encoding contexts as a history vector \( v_{his} \)
  - Semantic Label: ground-truth intent tags are encoded as the 1-hot sentence semantics
    \( \tilde{v}_{his} = BLSTM(\text{intent}_t) \)
  - Natural Language: CNN-encoded sentence vector for practical situations
    \( \tilde{v}_{his} = BLSTM(CNN(\text{utt}_t)) \)
  - 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

Speaker Role Modeling
- train two role-specific models independently, BLSTM\(_{rolea}\) and BLSTM\(_{roleb}\)
  \( \tilde{v}_{his} = BLSTM_{rolea}(\text{intent}_t, \text{role}_a) + BLSTM_{roleb}(\text{intent}_t, \text{role}_b) \)
  \( \tilde{v}_{his} = BLSTM_{rolea}(CNN(\text{utt}_t, \text{role}_a)) + BLSTM_{roleb}(CNN(\text{utt}_t, \text{role}_b)) \)

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

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

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

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