The model achieves language understanding for human-human dialogues between tourists and guides (DSTC4).

**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

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

- **End-to-End Training for Contextual Language Understanding**
  - BLSTM-encoded current utterance concatenated with the history vector for multi-label intent prediction
  
  \[
  \hat{y} = \text{BLSTM}(\vec{v}_{his}, \vec{x}) \quad P(\hat{y} | \vec{x}) = \prod_i P(y_i | w_1, \cdots, w_i)
  \]

  - All encoders, prediction models, and attention weights (except time-aware attention) can be automatically learned in an end-to-end manner

  \[
  \vec{v}_{his} = \sum_i \alpha_{\text{role}_i} \cdot \vec{v}_{his,\text{role}_i}
  \]

- **Attention Mechanism**
  - **Attention Type**
    - Content-Aware $\alpha^C$
    - Time-Aware $\alpha^T$
  - **Attention Level**
    - Sentence $\alpha_{ui}$
    - Role $\alpha_{ri}$

- **Recent Utterance contains more relevant information**
  
  \[
  \alpha^T_u = \frac{1}{d(u)}
  \]

  where $d(u)$ is the distance between $u$ and the current utterance.

- **Content-Aware Attention**
  - Semantic relation decides where and how much the model should focus on given the contexts
  
  \[
  \alpha^C_u = \text{softmax}(M_1(\vec{v}_{cur} + \vec{v}_{u}))
  \]

- **Role-Level Attention**
  - Different speaker roles behave differently
  - Role-level attention is based on how much to address on different speaker roles' contexts
  
  \[
  \alpha^R_r = \text{softmax}(M_2(\vec{v}_{cur} + \vec{v}_{his,r}))
  \]

- **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**

**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

**Code Available:**

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

---

Three authors contribute this work equally.

- Chun Chen
- Yun-Nung (Vivian) Chen
- Chung Chi

**Contact Information**

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