05

Multimodal situation recognition for everyday conversations

Dialog processing techniques for reading the situation

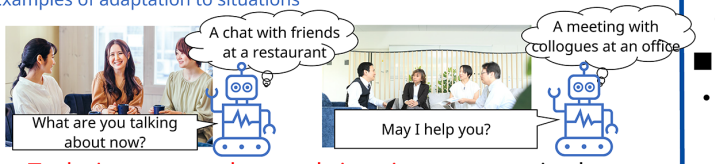
Abstract

Dialogue systems must correctly understand the dialogue situations to participate in everyday conversation naturally. We introduce two studies for understanding situations of everyday conversations. In the first research, we conducted a factor analysis to investigate the characteristics of everyday conversations. From the analysis, we clarified that our daily conversation could be explained by the combination of seven factors related to the dialogue purpose and manner. In the second research, we proposed a method to recognize dialogue situations, including the place of the conversation and the relationships among the participants by using multimodal information. In recent years, the naturalness of responses in dialogue systems has dramatically improved. Integrating multimodal information processing allows the system to have social conversations with humans in everyday situations. By developing our techniques, we aim to realize a future in which it will be commonplace to converse with dialogue systems.

Dialogue system for everyday conversations

■ To participate in everyday conversations, dialogue systems require the ability to adapt to situations

Examples of adaptation to situations



⇒ Techniques to understand situations are required

■ This presentation introduces two studies for understanding everyday conversation situations

1. Analysis and extraction of factors of everyday conversation
2. Dialogue situation recognition using multimodal information

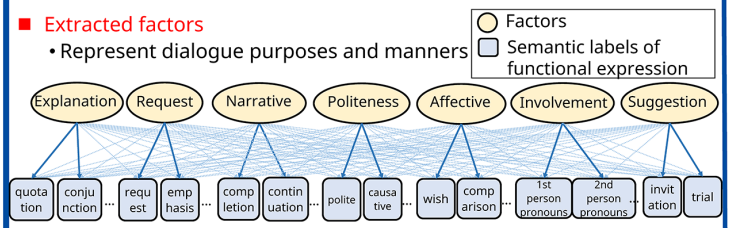
Analysis and extraction of factors of everyday conversation

■ Conducts a factor analysis of CEJC†

■ Features: semantic labels of functional expression
- Cues for capturing the characteristics of conversation
- Expresses modality, subjective information, tense, and so on

■ Extracts seven factors composing everyday conversations

■ Extracted factors
- Represent dialogue purposes and manners

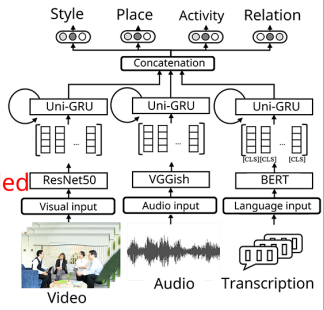


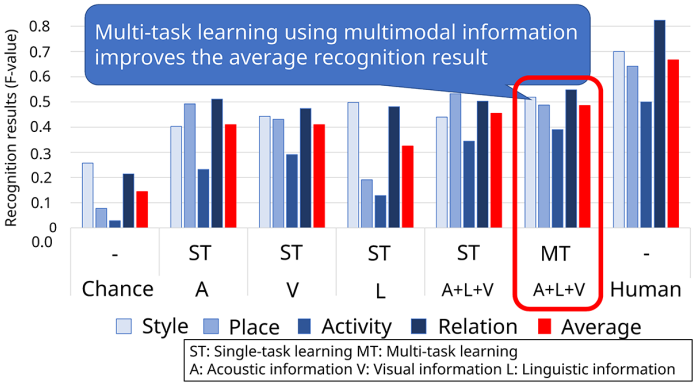
Multimodal dialogue situation recognition

■ Problem setting
- Recognizes situations incrementally by observing short scenes
- Uses acoustic, visual, and linguistic information

■ Recognition model
- Target situations
  - Style (e.g., meeting and chat)
  - Place (e.g., restaurant and car)
  - Activity (e.g., meals and rest)
  - Relation (e.g., family and friends)
- Multimodal inputs are embedded by pre-trained models
- Trained by multi-task learning considering relationships between situations

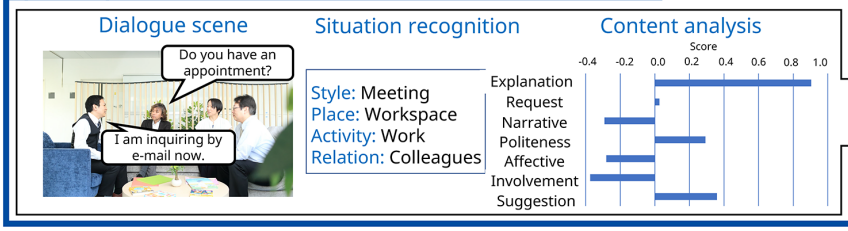
Overview of recognition model





Dialogue processing using proposed techniques (demo)

- Processes dialogue scene by using two methods of understanding dialogue situations
- Can adapt system response generation to situations and human-human conversation analysis



A part of the studies is based on joint research with Nagoya University.

†H. Koiso et al, Construction of the corpus of everyday Japanese conversation: An interim report, In Proc. LREC, pp. 4259-4264, 2018.

References

[1] Y. Chiba and R. Higashinaka, “Dialogue situation recognition for everyday conversation using multimodal information,” in *Proc. INTERSPEECH*, pp. 241-245, 2021.

[2] Y. Chiba and R. Higashinaka, “Analyzing variations of everyday Japanese conversations based on semantic labels of functional expressions,” *ACM Transactions on Asian and Low-Resource Language Information Processing*, Vol. 22, No. 2, pp. 1-26, 2022.

Contact

Yuya Chiba

Interaction Research Group, Innovative Communication Laboratory