

Title	内面状態の推定と対話戦略適応に基づく知的対話処理に関する研究
Author(s)	長澤, 史記
Citation	
Issue Date	2025-03
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/19928
Rights	
Description	Supervisor: 岡田 将吾, 先端科学技術研究科, 博士

Abstract

The ultimate goal of this research is to realize a computer system that can connect users to appropriate functions and services by empathizing with their feelings and interests through adaptive interactions and eliciting their inner needs. As a method for eliciting information from people, listening dialogue systems that use dialogues such as interviews and interviews have been actively studied. In listening dialogue systems, it is important to realize empathic dialogue, in which the system recognizes the inner state of the user and motivates the user to participate in the dialogue.

To realize such empathic dialogue processing, it is necessary to realize a mechanism in which the system proposes or changes topics according to the user's willingness, which is the user's inner state of "wanting to talk".

First, a machine learning model was developed to estimate the interviewee's willingness (desire to disclose information they have) based on their posture during the dialogue and the prosodic features of their speech utterances. Machine learning models using Random Forest and LinearSVM were trained. A method was developed to reduce the impact of individual differences in multivariate features on estimation accuracy for first-time interviewees who do not have the information necessary for normalizing multivariate features. Using the interview corpus collected through the dialogue experiment, this study evaluated the accuracy of the estimation of willingness by cross-validation, and found that the method correctly estimated high and low willingness with an accuracy of up to 72.8%.

Next, we introduced an adaptive dialogue strategy using this willingness recognition model and implemented it in a dialogue robot. In the adaptive dialogue strategy, if the estimated willingness is high, the robot continues the topic of the previous question, and if the estimated willingness is low, the robot switches the topic. Question selection by topic continuation/change was realized by exploring a pre-constructed question graph (a tree graph in which questions are arranged based on topic relevance). A dialogue experiment was conducted with 27 participants to evaluate the effect of adaptive dialogue strategy. The dialogue experiments were compared between the proposed adaptive dialogue strategy system and a random strategy system with random topic continuation/transition. The experimental results confirmed that the adaptive dialogue strategy gave users the impression that they were listening with more interest, and also significantly increased the number of

utterances with high willingness. This showed that even with less-than-perfect estimation accuracy, it is possible to motivate users to speak through adaptive dialogue strategy.

In order to improve the accuracy of multimodal inner state estimation and to analyze the accuracy of attitude estimation due to individual differences in multimodal features and the sensing environment, this study worked on refining the attitude estimation using the pre-built external corpus Hazumi1911. We trained and evaluated a model that added biodata and facial landmark features in addition to prosody and posture features. As a result, the addition of the features used improved the accuracy, and the individual differences in estimation accuracy decreased. We evaluated the accuracy of models trained on two different corpora with different sensing environments. The models were trained on the Hazumi1911 corpus and evaluated on the accuracy on data outside the corpus (a newly collected corpus of interview dialogues). The results showed that the accuracy of the models on data outside the corpus decreased, and the difference in accuracy between individuals also increased. However, the decrease in accuracy was smaller for models that used more features, and the difference in accuracy between individuals was also smaller.

A question generation method based on a large-scale language model (LLM) was proposed for the purpose of making adaptive dialogue strategy applicable to arbitrary topics. We implemented an improved interview robot system that incorporates updated willingness recognition model with extended features used and LLM-based adaptive question generation. The results of a dialogue experiment with 30 interviewees showed that the degree of self-disclosure of the interviewees improved when the adaptive strategy was used compared to the random strategy.

In summary, this thesis presents the results of an analysis of multimodal inner state estimation based on nonverbal information during dialogue, the implementation of an adaptive dialogue system, and its impact on dialogue. Our results show that an adaptive dialogue strategy increases user willingness, promotes self-disclosure, and lead to better interviews even with user adaptation using imperfect inner state estimation models.

These results will lead to new applications of dialogue technology through interview techniques that promote self-disclosure of the subject and elicit deeper narratives. For example, by eliciting the user’s unspoken feelings and inner narratives, it will be possible to improve counseling and service recommendations. The realization of such assistant technology that empathizes with the user’s inner world and proactively suggests solutions to the user’s problems will greatly improve future human-computer interaction.

Keywords: Sentiment Analysis; Physiological Signal Processing; Machine Learning; Multimodal Signal Processing; Dialogue System.