Title	ロバストな生体信号処理に基づくオンライン感性推定とマルチモーダル統合への応用
Author(s)	堅田，俊
Citation	
Issue Date	2022-12
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/18187
Rights	
Description	Supervisor:岡田 将吾，先端科学技術研究科，博士

Abstract

For modeling human intelligence, understanding emotional intelligence is an important and challenging issue. In affective computing, it has been reported that not only text, acoustic, and visual signals (observable signals) but also physiological signals (unobservable signals) are useful for estimating emotions and their related states. Physiological signals are expected to provide additional and less biased information compared with observable signals. Thus, coupled with growing interest in the development of emotionally intelligent systems, many studies related to physiological signals have been reported thus far; however, techniques that apply physiological signals for realistic emotion estimation tasks such as online (sequential) recognition for dialogue systems are still in the research phase, and there are unresolved issues in fundamental and applied research. In this thesis, three main research problems that have not previously been explored are addressed.

First, as one of the fundamental unresolved issues, physiological signals have individual differences that cause performance degradation of machine learning models based on physiological signals. Generally, it is assumed that both training and test data for machine learning are derived from the same distribution. Thus, estimation performance can degrade if there are physiological individual differences in unseen individual test data. In this thesis, physiological individual differences are considered a covariate shift to resolve this problem, and the Importance-Weighting (IW) method is introduced, which complements the model and is robust against individual differences for performance improvement of the models trained with physiological data. As a result, Importance-Weighted Support Vector Machine (IW-SVM) models outperform conventional models based on physiological features in emotion and personality estimation. These results indicate that IW in machine learning models can reduce the effects of physiological individual differences in physiological responses and contribute to the proposal of a new model for emotion and personality estimations based on physiological signals.

Second, although fundamental research on physiological signals provides insight into their potential, the effectiveness of physiological signals is often evaluated under emotion-evoked conditions. Thus, few studies have analyzed physiological signal effectiveness in naturalistic conditions. In particular, the evaluation and comparison of physiological signals with other observable signals under naturalistic human-agent interactions are insufficient. In human-agent interactions, it is necessary for the systems to identify the current internal state of the user to adapt their dialogue strategies. Nevertheless,

this task is challenging because the current user's sentiment is not always expressed by observable signals in a natural setting and changes dynamically. However, it is possible that physiological signals provide valuable information for online sentiment estimation since physiological responses cannot be consciously regulated. As applied research, a machine learning model based on physiological signals to estimate a user's sentiment at every exchange during a dialogue is presented in this thesis. Using a wearable sensing device, the physiological data including the Electrodermal Activity (EDA) and Heart Rate (HR) in addition to acoustic and visual information during a dialogue are evaluated. The sentiment labels are annotated by the participants (referred to as Self-repoted Sentiment (SS) label) for each exchange consisting of a pair of system and participant utterances. The experimental results show that a multimodal Deep Neural Network (DNN) model combined with the EDA and visual features achieves an accuracy of 63.2%. The analysis of the SS estimation results for each individual indicate that the human coders often incorrectly estimate negative sentiment labels, and in this case, the performance of the DNN model is higher than that of the human coders. These results indicate that physiological signals can help in detecting the implicit aspects of negative sentiments, which are acoustically/visually indistinguishable.

Finally, although the potential of the physiological signals in online SS estimation during dialogue is clarified in the abovementioned task, there is no comprehensive and thorough analysis of physiological signal application for multimodal fusion. Thus, the second task is extended by introducing different types of sentiment labels (annotated by third-party), which further clarify the contributions of physiological signals. Additionally, two state-of-the-art language models and six machine learning models, including recently reported multimodal DNN, are introduced. Furthermore, these analyses enable the creation of a robust multimodal physiological model that combines the proposed physiological signal processing method and the Transformer language model, named Time-series Physiological Transformer (TPTr). This model can capture sentiment changes based on both time-series linguistic and physiological information. In ensemble models, the proposed methods significantly outperform the previous best result ($p < 0.05$). These results provide new insight into machine learning methods that utilize both linguistic information and physiological responses during dialogue exchanges, which has not previously been explored.

In summary, this thesis presents novel robust physiological signal processing for emotion/sentiment estimation and its application to adaptive dialogue systems. This proposal will lead to a new application of physiological signals that are widely applicable in various fields. For example, the educational sys-

tem can capture the concentration level of students by monitoring students' internal states, and the psychological counseling system can be supported by understanding the context behind words. These emotionally intelligent systems will provide significant improvements in our lives in the future.

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