氏名	堅田 俊
学位の種類	博士（情報科学）
学位記番号	博情第 487 号
学位授与年月日	令和 4 年 12 月 23 日
論文題目	Robust Physiological Signal Processing for Online Sentiment Estimation and its Application for Multimodal Fusion
論文審査委員	主査 岡田 将吾 北陸先端科学技術大学院大学 准教授
	長谷川 忍 同 教授
	白井 清昭 同 准教授
	井之上 直也 同 准教授
	駒谷 和範 大阪大学 教授

論文の内容の要旨

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 system 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.

論文審査の結果の要旨

堅田氏は Affective Computing (感情計算論)の主要な課題の一つである，生体信号情報に基づく感情推定アルゴリズムに関する先駆的な研究を行った．近年，コミュニケーション中の人の感情状態を推定するために，発話言語（テキスト），音声，視覚信号（表情，ジェスチャ等）を統合したマルチモーダル情報に基づく機械学習が有効であることが明らかになっている．一方，感情状態は言語・音声・視覚といった観測可能な情報から推定できるとは限らず，本人の感情に関連して観測される生体活動の一部である心拍や皮膚電位を含む生体信号情報の処理は，今後有用となると期待されている．しかし，生体信号処理に基づくコミュニケーション中の人の感情状態の推定技術実現に向けて困難な点が多く残っていた．堅田氏は博士後期課程の研究で，上記の課題の解決に向けて， 3 つの研究を行った．

第一に，生体信号の個人差により，汎用的な感情推定のための特徴量抽出が難しく，機械学習モデルの精度低下の原因となる問題に焦点を当て，解決法を提案した．未知のテストデータと，訓練データの間に，生体信号特徴の個人差が含まれることに着目し，「訓練・テスト間で事前分布は異なるが感情推定の写像は共通であるという共変量シフト下での学習」の考え方を導入することで，個人差による精度低下を緩和する方法を提案・評価した．共変量シフト適応学習を導入することで，生体信号と感情・個人特性ラベルを含むベンチマークデータに対して，従来手法（2019 年時点での最高精度を示すモデル）の精度を改善できることを示し，個人差にロバストな推定が行えることを示した．

第二に，人と対話コミュニケーションを行う対話システムへの応用を目指し，発話毎にオンラインで感情状態を推定するモデルを提案・評価した．生体信号に基づく感情推定研究では，感情誘発条件下での実験で取得されたデータで評価を行っていることが多く，自然な対話コミュニケーション時に表出する感情の推定課題において，生体信号に基づくモデルはほとんど提案されていなかった．人・エージェントのマルチモーダル対話において，対話参加者自身がアノテーションした心象状態の推定タスクにおいて，生体信号特徴量に基づくマルチモーダル DNN (Deep Neural Network)を提案・評価した．皮膚電気活動 EDA と表情・ジェスチャの特徴を組み合わせたモデルは，63.2%の精度を達成した．また，複数名の対話参加者が心象を推定した結果は 63%であり，人間と同等程度の精度を得られたことが示された．

第三に，生体信号の役割をさらに明らかにすべく，複数の多面的感情ラベルと言語・音声・視覚・生体信号との関係を網羅的に分析し，本人の心象推定には，従来有効とされていた音声・表情等の情報が必ずしも有効でなく，生体信号情報と言語情報を組み合わせることが推定に有効であることを明らかにした．この結果に基づき，新規のアルゴリズムである，Timeseries Physiological Transformer (TPTr)を提案した．TPTr では，データ間の関係を自己注意（Self-Attention）機構により学習する Transformer 言語モデルに生体信号から抽出した時系列特徴量を統合し，言語・生体の時系列の間の関係を学習するために適している．TPTr モデルに基づき，生体信号のノイズにロバストな感情推定を実現できることを示した．

以上をまとめて，本論文では，感情推定のための頑健な生理的信号処理における問題点を明らかにし，生体信号に基づく感情推定のための新規の機械学習モデルを提案・評価した． 提案したモデルは人・システム対話においてオンラインに変化する感情状態を精緻に予測できることを示した．

以上，本論文は，マルチモーダルインタラクションにおける生体信号処理の課題・その解決方法を明らかにしたものであり，学術的に貢献するところが大きい．よって博士（情報科学）の学位論文として十分価値あるものと認めた．