

Title	人 - ロボット相互作用における人の性格特性推定に向けた選択的マルチモーダル融合アプローチ
Author(s)	申, 志豪
Citation	
Issue Date	2021-03
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/17477
Rights	
Description	Supervisor: 丁 洛榮, 先端科学技術研究科, 博士

氏 名	SHEN, Zhihao		
学 位 の 種 類	博士(情報科学)		
学 位 記 番 号	博情第 450 号		
学 位 授 与 年 月 日	令和 3 年 3 月 24 日		
論 文 題 目	A Selective Multi-modal Fusion Approach to Inferring Human Personality Traits in Human-Robot Interaction		
論 文 審 査 委 員	主査	CHONG, Nak Young	JAIST Professor
		NGUYEN, Minh Le	JAIST Professor
		OKADA, Shogo	JAIST Assoc. Professor
		SGORBISSA, Antonio	University of Genoa Assoc. Professor
		AHN, Ho Seok	Univ. of Auckland Senior Lecturer

論文の内容の要旨

With the population aging and sub-replacement fertility problems increasingly prominent, many countries have started promoting robotic technology for assisting people toward a better life. The robot was designed with the appearances that are similar to human's. And more importantly, many robots also were endowed with many capabilities such as synchronized verbal and nonverbal behaviors, emotion recognition, and many others for acquiring a high-quality interaction between robot and its users.

It has been found that the personality traits are playing very important roles in human-human interactions. With an increasing number of research on personality traits, their relationship to many important aspects of life, such as job performance, health-related behaviors, emotion, and many others have been revealed. Therefore, understanding personality traits is useful for predicting human behaviors, and understanding human's mind and how personality traits affect the attitude and behaviors towards other people. Once the robots are endowed with the capability of recognizing human personality traits, the robots then will be able to adjust their behaviors such as voice volume, speech rate, and body gestures to enhance the degree of user engagement.

For achieving this goal, a pilot experiment for personality traits recognition was conducted for testing the feasibility of inferring personality traits from nonverbal behavior features, and finding more practical problem in human-robot interaction. Some features which are head motion, gaze, body motion, voice pitch, voice energy, and Mel-Frequency Cepstral Coefficient were extracted to describe human's nonverbal behaviors. Each feature showed its advantage in a different aspect. However, different nonverbal features can provide different personality traits classification results. It is not a standard way of drawing the conclusion for declaring the user's personality traits. On the other hand, the camera was fixed to make sure that the background did not change, same strategy also was applied in many related studies for the same purpose. However, this conflict with the idea that robot that was enabled to understand human personality traits aims to behave more properly.

Therefore, a new paradigm of human-robot interaction as close to the real situation as possible was designed, the following three main problems were also addressed: (1) fusion of visual and audio features of human interaction modalities, (2) integration of variable length feature vectors, and (3) compensation of shaky camera motion caused by movements of the robot's communicative gesture. Same nonverbal features including head motion, gaze, and body motion, voice pitch, voice energy, and Mel-Frequency Cepstral Coefficient were extracted from a camera mounted on the robot performing verbal and body gestures during the interaction. Then, the system was geared to fuse these feature and deal with variable length multiple feature vectors. Lastly, considering unknown patterns and sequential characteristics of human communicative behavior, a multi-layer Hidden Markov Model that improved the classification accuracy of personality traits and offered notable advantages of fusing the multiple features was proposed. The results were thoroughly analyzed and supported by psychological studies. The proposed multimodal fusion approach is expected to deepen the communicative competence of social robots interacting with humans from different cultures and backgrounds.

Keywords: Human-Robot Interaction; Personality Traits Recognition; Multimodal Feature Fusion; Nonverbal Features; Multi-layer Hidden Markov Model; Machine Learning Model.

論文審査の結果の要旨

This dissertation addresses the problem of probabilistic inference of human personality traits during human-robot social interaction using the user's non-verbal communication cues recognized by social robots equipped with on-board sensors. Given the reality of social interaction in everyday life that occurs regardless of location, the robot's first-person perspective is used to extract visual features such as the user's head motion, gaze, and body motion, and audio features such as vocal pitch, vocal energy, and mel-frequency cepstral coefficient, without using any environment-embedded sensors. Along the lines, the author proposed an efficient fusion technique of the aforementioned visual and audio features of human interaction modalities, integrating variable length feature vectors. He furthermore compensated for shaky camera motion caused by the movements of the robot's communicative gesture. The effectiveness of the proposed approach was verified through several different human-robot communication scenarios with participants in the real world environment settings.

Specifically, the author implemented a multi-layer Hidden Markov Model on an off-the-shelf social robot that significantly improved the classification accuracy of the user's personality traits taking advantage of selective feature fusion and machine learning classifiers. The results were thoroughly analyzed and compared to the psychological nature of human beings. The proposed computational models have been published in the IEEE/CAA Journal of Automatic Sinica (IF 5.129) and several IEEE conference proceedings, and the author won the Best Paper Award in Advanced Robotics of the 2019 IEEE International Conference on Advanced Robotics and Mechatronics. In contrast to most of existing approaches,

considering unknown patterns and sequential characteristics of human communicative behavior, this work explores automatic selection of important/influential features and offers a comparative analysis of different feature combinations. Overall, this research opens doors to impactful contributions to the social intelligence and human-robot interaction community. This is an excellent dissertation and we approve awarding a doctoral degree to SHEN Zhihao.