Title	Inferring Human Personality Traits in Human-Robot Social Interaction
Author(s)	Shen, Zhihao; Elibol, Armagan; Chong, Nak Young
Citation	2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI): 578-579
Issue Date	2019-03-11
Type	Conference Paper
Text version	author
URL	http://hdl.handle.net/10119/16107
Rights	This is the author's version of the work. Copyright (C) 2019 IEEE. 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 2019, pp.578-579, DOI:10.1109/HRI.2019.8673124. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
Description	

Inferring Human Personality Traits in Human-Robot Social Interaction*

Zhihao Shen, Armagan Elibol, and Nak Young Chong
School of Information Science
Japan Advanced Institute of Science and Technology
Nomi, Ishikawa 923-1292, Japan
{shenzhihao, aelibol, nakyoung}@jaist.ac.jp

***Abstract*—In this report, a new framework is proposed for inferring the user's personality traits based on their habitual behaviors during face-to-face human-robot interactions, aiming to improve the quality of human-robot interactions. The proposed framework enables the robot to extract the person's visual features such as gaze, head and body motion, and vocal features such as pitch, energy, and Mel-Frequency Cepstral Coefficient (MFCC) during the conversation that is lead by Robot posing a series of questions to each participant. The participants are expected to answer each of the questions with their habitual behaviors. Each participant's personality traits can be assessed with a questionnaire. Then, all data will be used to train the regression or classification model for inferring the user's personality traits.**

***Index Terms*—human-robot interaction, social cue, user personality traits, regression model, classification model**

I. Introduction

Nowadays domestic service robots are becoming more and more common in our daily life. They play important roles in various aspects such as accompanying older people, taking care of patients, entertainment, and several others. Synchronized verbal and non-verbal behaviors have been made possible in various humanoid robot platforms, such as ASIMO, Nao, Pepper, to allow them to interact better with people. Those behaviors were designed to make the robots draw people's attention and maintain effective engagement with them. However, there is a clear need for strategies to enable the robot to respond to the user's behaviors appropriately.

There are several studies that are focusing on improving the performance of human-robot interaction by enabling a robot to interact with humans in a natural manner and to understand the user's behaviors, feelings, and even thoughts. This extremely arduous task requires that the robot should be able to understand the user's actions and internal states so that it will be able to respond to the user's intention. In the course of human-human interaction, humans can comprehend the personality traits of the person they are interacting with from the person's behavioural responses. Furthermore, humans can adapt their behaviours to enhance the engagement. It has been showed that there is a strong connection between personalities and behaviours in human-robot interaction [1], [2]. There are some research efforts on another interesting concept about the complementarity attraction [3], [4], which means that an individual enjoys talking with someone even more whose personalities are complementary to his or her personalities. In light of these, inferring the user's personality traits is one of the most important capabilities which social robots should be endowed with to increase user engagement and comfort.

*This work was supported by the EU-Japan coordinated R&D project on "Culture Aware Robots and Environmental Sensor Systems for Elderly Support," commissioned by the Ministry of Internal Affairs and Communications of Japan and EC Horizon 2020 Research and Innovation Programme under grant agreement No. 737858.

II. Feature Representation

We use a social humanoid robot Pepper which was created to make people happy, enjoy their life, facilitate the relationship, and connect people with the outside world. Each participant will have little chitchat with the robot about the weather, food, physical condition and etc.

A. Methodology

For designing and training our model, the procedure is briefly explained in the Fig. 1. Step a), the user is asked to fill out a questionnaire [5] for assessment of their personality traits; Step b), during the face-to-face interaction, Pepper records video features like the user's head motion, gaze score, and motion energy as well as vocal features like pitch, energy, and MFCC; and Step c), we design our model with the machine learning technique, and train the model with feature data from Step b and personality trait labels from Step a.

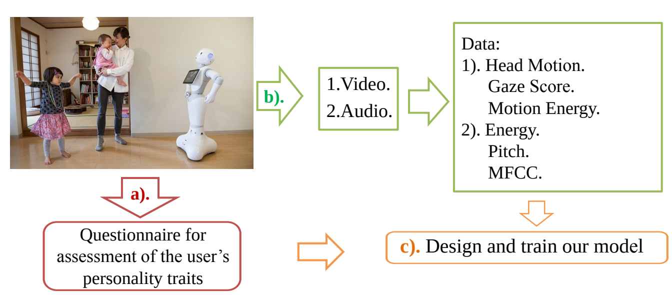

Fig. 1. Inferring Personality Traits in Human-Robot Interaction

B. Personality Traits

We are going to measure the user's the Big-Five Personality Traits which has been adopted in psychology as common descriptors of personality in the most of existing researches on personalities. The five broad personality traits can be described as Extroversion, Agreeableness, Conscientiousness, Emotional Stability, and Openness to experience.

C. Non-verbal Features

The non-verbal features are defined as three visual features including the user's head Motion, gaze score, motion energy, and three vocal features such as pitch, energy, and MFCC.

1) Head Motion

The rotation invariant multi-view face detection method [6] is used to detect the user's face from the image that was captured by the robot's camera. Once the face position is detected, the sub-window that contains a human face is used for pose estimation. The head pose is the 3D orientation (yaw, pitch, and roll) of the detected face. The Head Motion is computed as the Manhattan distance of two adjacent head poses.

2) Gaze Score

We consider that the gaze direction is related to the face plane. Therefore, it is possible to detect whether the person is looking at the robot or not by analyzing the head angles. The gaze score (between 0 and 1) tells whether the person is looking at the robot or not.

3) Motion Energy

The motion energy is another important visual feature. It can be acquired from a long period of time over the whole human-robot interaction. We calculate the proportion of the moving pixels relative to the total number of pixels in the current frame. Each frame is compared with the previous one. The different pixels are identified as moving pixels. It is easy to calculate the moving pixels for all the points in the frame based on the dense optical flow [8] when the camera is fixed. However, we prefer that the robot is talking with gestures, which means the camera is movable. In the pinhole camera model, any two images of the same planar surface are related by a homography. Therefore, wthe best-matched keypoints in both images can be extracted by using SIFT. If enough matched keypoints are found, these keypoints are passed to calculate the transformation matrix. A perspective transformation is used to project the current image point into the previous image plane to verify whether there is a change.

4) Pitch and Energy

Pitch is interpreted as the frequency of the sound by ear and brain. In this paper, we use auto-correlation function to calculate the short term pitch sequences in the time domain [7]. In the following equation, $acf(\tau)$ is the auto-correlation function, $s(i)$ is the audio signal of a frame, N is the frame length, τ is a delay:

$$acf(\tau) = \sum_{i=1}^{N-1-\tau} s(i)s(i+\tau), (0 \leq \tau < N) \qquad (1)$$

The pitch of each frame is calculated by dividing the sampling frequency by the second peak location of auto-correlation functions (the first peak is when τ is 0). Then, we slide the window to the next frame until the end of the audio signal.

We calculate the normalized energy of each frame by using the following equation, where $s(i)$ is the audio signal of a frame, N is the frame length.

$$Energy = \frac{1}{N}\sum_{i=1}^{N} s(i)^2 \qquad (2)$$

5) Mel-Frequency Cepstral Coefficient

The frequency of the incoming sound can vibrate different spots of human cochlea. Depending on the different location in the cochlea, different nerves were stimulated to inform the brain that some frequencies are present. MFCC was proposed based on this concept as it is close to what humans actually hear. We investigate how these essential features affect people's perception of other people's personality traits. The method that was proposed in [9] is used to calculate MFCC.

III. Experimental Design

We assume that each participant is standing or sitting 1.5 to 1.7 meters in front of the robot. In the scenarios, the participant mainly is talking with the robot during the interaction. Therefore, the conversation is lead by the robot by posing questions which are related to some key-words from the participant's answers. The non-verbal features and personality traits of each participant will be used to train a machine learning model such as SVM, ridge regression, and etc [10]. Then, their performance of the classification and regression models will be evaluated in the final step.

References

[1] S.N. Woods, K. Dautenhahn, C. Kaouri, R.T. Boekhorst, K.L. Koay, and M.L. Walters. Are robots like people? Relationships between participant and robot personality traits in human-robot interaction studies. Interaction Studies, 8(2):281–305, 2007. 2

[2] C. Nass and M.K. Lee. Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency attraction. Experimental Psychology: Applied, 7:171–181, 2001. 2, 4

[3] K. Isbister and C. Nass. Consistency of personality in interactive characters: Verbal cues, non-verbal cues, and user characteristics. Human-Computer Studies, 53:251–267, 2000.

[4] T.F. Leary. Interpersonal diagnosis of personality. Ronald Press, NY, USA, 1957.

[5] Goldberg, Lewis R. The development of markers for the Big-Five factor structure. Psychological Assessment 4.1 (1992): 26.

[6] B. Wu, H. Ai, C. Huang and S. Lao, "Fast Rotation Invariant Multi-View Face Detection Based on Real Adaboost", IEEE Conf. on Automatic Face and Gesture Recognition, pp. 79- 84, 2004 .

[7] P. Boersma. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. in Proc. of the Institute of Phonetic Sciences, vol. 17, no. 1193. Amsterdam, 1993, pp. 97–110.

[8] Farnebäck, G. Two-frame motion estimation based on polynomial expansion. In Proceedings of the Scandinavian conference on image analysis. 2003

[9] Min Xu; et al. (2004). "HMM-based audio keyword generation". In Kiyoharu Aizawa; Yuichi Nakamura; Shin'ichi Satoh. Advances in Multimedia Information Processing – PCM 2004: 5th Pacific Rim Conference on Multimedia (PDF). Springer. ISBN 3-540-23985-5. Archived from the original (PDF) on 2007-05-10.

[10] Cigdem Beyan, Francesca Capozzi, Cristina Becchio, and Vittorio Murino. Prediction of the Leadership Style of an Emergent Leader Using Audio and Visual Nonverbal Features. IEEE Transactions on Multimedia 2017.