Expert Systems with Applications 42 (2015) 4540–4555

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

A novel multimodal communication framework using robot partner for aging population

Dalai Tang [a], Bakhtiar Yusuf [a], János Botzheim [a,b,*], Naoyuki Kubota [a], Chee Seng Chan [c]

[a] Graduate School of System Design, Tokyo Metropolitan University, 6-6 Asahigaoka, Hino, Tokyo 191-0065, Japan
[b] Department of Automation, Széchenyi István University, 1 Egyetem tér, Győr 9026, Hungary
[c] Centre of Image and Signal Processing, Faculty of Computer Science & Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia

ARTICLE INFO

Article history:
Available online 19 January 2015

Keywords:
Human–robot interaction
Robot partner
Intelligent robot
Aging
Gesture recognition

ABSTRACT

In developed country such as Japan, aging has become a serious issue, as there is a disproportionate increasing of elderly population who are no longer able to look after themselves. In order to tackle this issue, we introduce human-friendly robot partner to support the elderly people in their daily life. However, to realize this, it is essential for the robot partner to be able to have a natural communication with the human. This paper proposes a new communication framework between the human and robot partner based on relevance theory as the basis knowledge. The relevance theory is implemented to build mutual cognitive environment between the human and the robot partner, namely as the informationally structured space (ISS). Inside the ISS, robot partner employs both verbal as well as non-verbal communication to understand human. For the verbal communication, Rasmussen's behavior model is implemented as the basis for the conversational system. While for the non-verbal communication, environmental and human state data along with gesture recognition are utilized. These data are used as the perceptual input to compute the robot partner's emotion. Experimental results have shown the effectiveness of our proposed communication framework in establishing natural communication between the human and the robot partner.

1. Introduction

From the year of 2000 to 2050, it is expected that the proportion of elderly people (60 years old or older people) in the world's population will double from 11% to 22%, and the absolute number of elderly people is expected to increase from 605 million to 2 billion over the same period.[1] Meanwhile, in Japan, according to the Statistics Bureau at the Ministry of Internal Affairs and Communication,[2] the population of elderly people is expected to increase to 36 million, that is about 31% of the population in the year of 2030. In Tokyo itself, it is anticipated that the number of elderly people will reach 25.2% of the population in year 2015.

Along with the increasing number of elderly people, one must note that the number of those elderly people who are no longer able to look after themselves will also increase proportionally. Many of them will lose the ability to live independently because of limited mobility, frailty or other physical or mental health problems (Chernbumroong, Cang, Atkins, & Yu, 2013; Rueangsirarak, Atkins, Sharp, Chakpitak, & Meksamoot, 2012). In Japan, the increasing number of elderly people who live alone or independently has required a large number of nursing care to support them. However, since the number of caregivers is always limited, it is important to introduce alternative solution to tackle this problem. One of the solutions is the introduction of the human-friendly robot partner to support the elderly people in their daily life.

According to Broekens, Heerink, and Rosendal (2009), there are two types of robots that are able to support the elderly people. One is the rehabilitation robot and the other is the social robot. In the former, the robot focuses on physical assistance technology, whilst the latter is concerned as a system that has the capability in human–robot communication. It is also known as the robot partner. This paper focuses on the latter with a focus on realizing a natural communication between the human and the robot partner. The natural communication can be realized when robot can understand human intention or thought. We implemented the theory of relevance (Sperber & Wilson, 1995) to build mutual cognitive environment between human and robot partner into our system, which

* Corresponding author at: Graduate School of System Design, Tokyo Metropolitan University, 6-6 Asahigaoka, Hino, Tokyo 191-0065, Japan. Tel.: +81 42 585 8441.
E-mail addresses: tang@ed.tmu.ac.jp (D. Tang), yusuf-bakhtiar1@ed.tmu.ac.jp (B. Yusuf), botzheim@tmu.ac.jp, botzheim@sze.hu (J. Botzheim), kubota@tmu.ac.jp (N. Kubota), cs.chan@um.edu.my (C.S. Chan).

[1] See www.who.int/ageing/en/.
[2] See www.stat.go.jp/english/data/handbook/c0117.htm.

http://dx.doi.org/10.1016/j.eswa.2015.01.016
0957-4174/

called informationally structured space (ISS) to handle this problem. According to Sperber and Wilson (1995), relevance theory is very useful to discuss the multimodal communication, where each person has his or her own cognitive environment that make their communication restricted. Therefore, usually humans use their utterances or gestures to expand their cognitive environment by extracting person's attention into specific target object, event, or person. When human's cognitive environment became wider, they can share each other intention or thought as illustrated in Fig. 1. The implementation of this theory into our system can be observed in the structure of database in our system, which is explained more detailed in Section 6. Meanwhile in conducting communication between human and robot, we use the Rasmussen's behavior model to build the conversation system. In addition to verbal communication, we also implement non-verbal communication such as facial expression, emotional gestures and pointing gestures.

Our contribution of this paper is to treat all these elements (environment recognition, human recognition and emotional model) as an unified framework in the informationally structured space, so that a more natural communication between a robot partner and a person can be formed. In order to facilitate this, we built a new type of robot partner, named "iPhonoid". Experiments using three different case studies have shown the effectiveness of the proposed framework in establishing natural communication between the human and the robot.

The rest of the paper is organized as follows. Section 2 discusses the literature related to the proposed system. Here, we also explain the advantage and disadvantage of the proposed method compared to previous researches. Section 3 introduces the concept of informationally structured space. Section 4 deals with the environmental system, which includes sensor network, web system and robot system, while robot system is explained separately in the following section. Section 5 explains the robot partner including gesture recognition technique and emotional model. Section 6 discusses the database system and communication system. Section 7 details the conversation system. Section 8 presents experimental results of the proposed method. Section 9 summarizes and discusses the future direction of this research.

2. Literature review

In the proposed method, we implemented relevance theory to build mutual cognitive environment called informationally structured space; Rasmussen's behavior model for conversational system; emotional model and gesture recognition to realize natural communication between human and robot partner. In this section

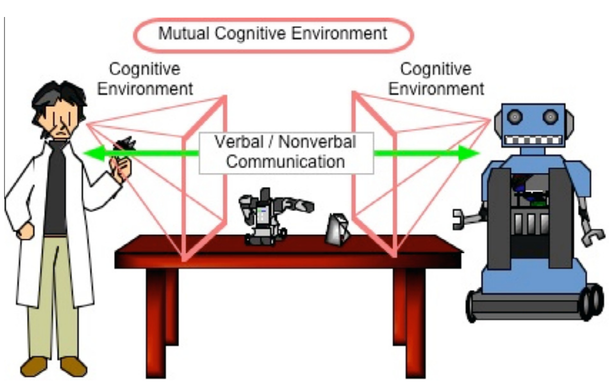

Fig. 1. Mutual cognitive environments via natural communication. In order to realize a natural communication with a person, the robot partner acquires information on the surrounding environment, as well the people's condition by processing the data obtained from various sensors. When the robot partner is able to build a mutual cognitive environment, the robot partner can understand the people's intention or thought.

we discuss and compare our proposed method to previous researches. Since mutual cognitive environment has a close relationship with ambient intelligence, we will start to review previous researches on this field. Next, we will discuss emotional model, thereafter gesture recognition and finally conversation system.

In works related to ambient intelligence Montes, Ortega, Venzala, and Abril (2014) built smart environment based on software reference architecture. The smart environment is used to conduct the perception process in a standard office. However, the paper only used motion detection in order to measure data from sensor. Details such as gestures were not included. On the other hand, Lee, Lee, Kim, Wang, and Love (2014) proposed a method called mixed context-aware inference, which is a novel sensor-based context-aware system focusing on three inference processes: rule, inference and pattern driven. Forkan, Khalil, and Tari (2014) used various sensors to measure data, which enabled this method to get more accurate result. Moreover, the usage of cloud technology made the process time become shorter. However, this method is difficult to realize concerning the high cost. Furthermore, since user has to wear special clothes to get data from the sensors, it is very cumbersome. Forkan, Khalil, Tari, Foufou, and Bouras (2015) proposed fusion-based architecture, detection in activity and location patterns using Hidden Markov Model (HMM). Although it has a good accuracy in computation, HMM has a disadvantage in high computation cost.

For the emotional model concept, Tay, Jung, and Park (2014) proposed a method, which effects on occupational roles (security vs. health-care), gender (male vs. female), and personality (extrovert vs. introvert) on user acceptance of a social robot. However, they only use stereotype to conduct the evaluation, which makes the result arguable. Daosodsai and Maneewarn (2013) proposed a method to generate emotion of a robot using expert knowledge by fuzzy logic. The emotion of the robot is determined using 3 types of input data, such as the robot's personality, the ambient environment, and the interaction with human. However, since emotional expression is using LED only, the emotional expression done by the robot has a lot of limitations and it is difficult to be evaluated. Emotional model proposed by Kim, Yang, and Kwon (2013) used episodic memory system, as a result of long term human robot interaction and emotion generation reaction. Although the method of this paper is very interesting, the application is only possible in the virtual environment. Jitviriya and Hayashi (2014) used the integration of environment, robot self-states and feedback behavior for generating robot emotion (human data is not included). As a conclusion in the previous research the definition of the robot's emotion is not clear, while in our paper the change in environment, human state (gesture and distance) effects the robot's emotion, which linked into the conversation system content and robot's facial expression.

For the works related to gesture recognition, Iengo, Rossi, Staffa, and Finzi (2014) proposed a novel approach to real-time and continuous gesture recognition for flexible, natural, and robust human–robot interaction (HRI), and the generation of an ad hoc HMM. As mentioned before, one disadvantage of HMM is the high computational cost. Gesture recognition method proposed in Xiao, Yuan, and Thalmann (2013) is based on combination of the CyberGlove and Kinect sensor, which could recognize various gestures. The using of the CyberGlove for measurement device shows that this method cannot be directly realized in daily life now owing to its price. Liu, Hu, Luo, and Wu (2014) proposed a method in gesture recognition by using on-board monocular camera and specialized gesture detection algorithms. Here, also the dynamic movement primitives (DMP) model is employed. In this method, since the depth information is not acquired, the recognition has many limitations. The gesture recognition in our system used Kinect sensor has lower the cost, although our method is not as