Review of Human-Robot Interactive Modelling and Application for Elders

Jing Han, Lun Xie*, Shangmou Xu, Zhiliang Wang

School of Computer and Communication Engineering, University of Science and Technology Beijing, No.30 Xueyuan road, Beijing, China

Received 23 November 2014, www.cmnt.lv

Abstract

As the psychological problems of the aged are increasingly obvious in the aging society, the human-robot interaction technology making psychological adjustment for the aged becomes one of the most important directions of robotics’ research. Firstly, this paper describes the current development situation of interaction technology that can be applied for the service robot of elderly psychological adjustment. Next the methods of feature recognition and sentiment analysis in facial expression interaction, speech interaction, and other interactions are discussed. Then several typical emotion models are analyzed in detail. Finally, the possible research directions of robot interaction technology for elderly psychological adjustment are summarized and prospected.

Keywords: Human-robot interaction; Expressive interaction; Speech interaction; Emotion model; The service robot of elderly psychological adjustment.

1 Introduction

A great number of age-related diseases, transformation of social roles, decrease of their somatic functions and changes of family structures have strong impacts on the old people’s quality of life and mental states [1]. With an increasing number of aging population, the problem of elderly mental health has been appearing gradually, which leads to a big challenge to states and individuals. Under the rapid development of intelligent robotic industry, it’s increasingly necessary to solve the problem of an aging society by using the robot. A robot of elderly psychological adjustment acts as a new type of “company” into sight and sound. Intelligent human-robot interaction is an important premise and a key technology of the elderly mental health service robot. In the process of interaction, the service robots can analyze interactive information of the old through affective computing models and perceive psychological states of the old [2-3]. The service robots have a great deal of abilities of emotional recognition and emotional feedback communicating friendly with the elder people. Human-robot interaction is supported by cameras, microphone and Kinect, etc. utilizing facilities to sample emotional information of the elder people and identifying mental states by a series of algorithms. Robotic interactions have the features of reliability, stability and friendliness.

This paper describes a series of typical human-robot interactions of the elder people’s mental health service robots. The second part shows its development of dynamic state. More interactive methods of robot and process of the different interactions are discussed in the third part, such as facial expression interactions, voice interactions, etc. It is further discussed the algorithms and models of the emotion analysis in the third part and fourth part. In the last part, this paper summarized and prospected the interactive technologies in the robot of elderly psychological adjustment.

2 Research and development

The elderly psychological health robots can not only communicate with old people, but decide the psychological changes from their facial features, gesture and others information having the ability of psychological adjustment 4. It’s presented that Professor Minsky, M.I.T. He believed that the intelligent machines could not be developed without emotion [4]. And then under the trend of the aged society, the governments and domestic and foreign experts pay attentions to the research of the interactions of old people’s psychological health robots.

Service robots and other high technologies to assist pension have been taken for many years in the USA. National Science Foundation (NSF) gives a great deal of financial support to the research projects annually. The human-robot interactive technologies have been applied to varieties of the elderly health service robots successfully. The Kismet robot developed by M.I.T. has capabilities of baby-like behaviours and interactions. It has 15 degrees of freedom in its face and can mimic varieties of human expressions. Its visual and auditory functions consist of the camera in its eyes and the microphone in its ears. This emotional model of the Kismet robot converge external input and internal needs with behaviour actions communicating with people fluently.

According to the national strategy of a five-year technological plan from “e-Japan” to “u-Japan” announced in 2000, Japan, the service robots have been becoming the main research directions to solve the problem of the aged society in Japan. The aging health robotic interactive research has made great progress. For example, Paro, the robot for the elderly psychological healthy service, makes full use of interactive information of facial expression, voice, postures and physical characteristics, via analyzing the olds’ psychological and emotional states from external input. Many research institutions launched a series of

* Corresponding author’s e-mail: xielun@ustb.edu.cn
trials’ analysis to robot Paro. The Robot Paro has a good
effect of adjuvant therapy to elderly mental illness from the
test data [5-6]. Additionally, the “Hector” robot developed
by the EU’s Seventh Framework Program (Fp7) makes full
use of the kinematic analysis, facial recognition and voice
recognition to capture the facial features, even though users
are drinking and reading newspapers. It is easy to achieve
the process of natural interactions. Friendly interactive
technology is one of the important features [7-8].

Many Chinese experts study relative technologies of
the service robots for the aged people. Robot “Huihui” in-
vented by the Chinese Academy of Sciences utilizes emo-
tional speech database and affective computing models. It
is able to communicate with the elderly people to under-
stand their psychological states [9]. In addition, there are
many universities and research institutions doing resear-
ches for the robotic interactive techniques deeply. Due to
immature technologies of wireless sensor networks, face
and voice recognition, affective computing model, interac-
tions of robots are still developing and there is a gap to
achieve full and friendly, natural, safety human-robot
interactive environment. It is required to do abundant ex-
periments improving the robotic performances and indicators
for the natural interaction.

3 Human-robot interaction in elders

Friendly and natural human-robot interactive environment
reflects the service robots’ intelligence providing more
personalized services and psychological adjustment for
elders. Considering the old man’s physical condition, hu-
man-robot interaction in the elderly need to have the follo-
ing features:
• Harmony means of interaction
• Effective emotion analysis and modelling
• The high fusion of interactive information and emotion
  model
• Simple, easy to use and nice human-robot interface
• Powerful information storage and real-time data
  processing

Albert Mehrabian, an American psychologist, dis-
covered that 7% affective information is from language, 38% is
from tone, and 55% is from facial expressions and body
languages respectively [10]. It’s apparently shown that the
voice, expressions and actions analysis are better to reflect
the emotional states of the olds. By means of learning the
way of human beings, experts have done lots of research in
the field of face interactions, voice interactions, and word
interactions. Moreover, because the analytic process of
human cognitive information is complex during the period
of communication, the service robots’ learning algorithms
of interactive information are multi-modal and complicated

In the progress of human-robot interaction, the service
robots sample and process the users’ information by a
series of facilities. The sampling information consists of
facial expressions, voice, physical characteristics index,
gestures and words, etc. Next, this information is needed
for feature extraction and model analysis by the service
robots. Furthermore, according to the classified informa-
tion and emotional states, the service robots make mat-
ching analysis and emotion regulation. Last, the service
robots express the corresponding output on the basis of
emotion analysis results. The specific interaction process is
shown in Figure 1.

![Figure 1: The flow chart of the service robots’ human-robot interaction](image)

In daily life, we can get the other person’s emotional
state by eyesight, face, voice and posture and make a
comprehensive analysis based on access to a variety of
expression modes. The service robots, by means of such a
series of hardware equipment as cameras, microphones,
sensors, etc. to capture the interactive information, use the
information detection and recognition technology to deter-
mine interactive emotional state. There are many methods
can be used in feature extraction and recognition of interac-
tive information. Some algorithms, for example, Principal
Component Analysis (PCA), Gaussian Mixture Model
(GMM), Hidden Markov Model (HMM) and Finite State
Machine (FSM), can be used in a variety of interactive
information.

3.1 EXPRESSIVE INTERACTION

Recently, facial expression recognition, as one of the hot
research topics in human-robot interaction, can bring the
elderly people face to face communication. The mental
health status of elderly people will be harvested into the
robot from video chats.

In the process of facial expression recognition, face
feature information is detected and extracted, and then it is
classified and trained by the reasonable technologies. The
typical process for a facial expression recognition system is
shown in Figure 2. Generally there are two kinds of algo-
rithms for facial expression recognition: one is a face
recognition algorithm based on geometric feature, the other
is based on appearance characteristics [12]. The former
method takes into account some facial geometric features
and its relationship as classification feature information,
including eyes, mouth and nose, etc; In 1971, based on
human facial expression changes, Ekman and Friesen have
put forward six basic emotion classifications: surprise, fear,
disgust, anger, happiness, sadness [13]. According to the
changes of the facial muscles in the expression occurrence,
the Facial Action Coding System (FACS) was proposed by
Ekman and Friesen in 1978. It describes the corresponding
relationship between Action Unit (AU) and facial expres-
sions, and has greatly promoted the progress facial expres-
sion recognition research; Facial features detection usually
adopts methods such as template matching, motion extrac-
tion gray histogram [14]. Facial expression detection can be

409
divided into static image and video sequence according to different input mode.

The latter method analyzes the facial expression by means of the texture feature of whole facial area and characteristic distribution in high dimensional space. Jennifer Huang [15] constructed 3D Morphable Models based on the variable of input face image. The system overcame the rigid requirements from the illumination to the outside and invariant pose in the interaction process, whose recognition rate was 90% in the database owing 1200 facial images. In addition, due to the interference of the external environment such as illumination, object shade seriously affect the facial expression recognition rate; Mahdi Ilbeygi [16] proposed a system based on fuzzy reasoning model to reconstruct the image in 2012, which could detect the obscured facial expression changes and set up the recognition model that enables to identify six kinds basic facial expressions used the optimal fuzzy membership functions deduced by the genetic algorithm. The average recognition rate was not less than 91.3% in multiple databases of different faces.

3.2 SPEECH INTERACTION

Language is one of the main ways of daily communication. After analysis the spoken information, people use language to express their feelings. In the research of elderly health service robot, voice is the simplest and most effective means of emotional communication. Through the analysis of semantic, tone and rhythm determine, the emotional state of communicators can be determined. For a long time, the development of speech recognition systems in the field of human-robot interaction has made an indispensable contribution. Currently, voice interaction recognition rate can reach more than 95%under the quiet environment. During computer operation, the speech can be better than that of the keyboard and mouse interaction effects [17].

Audio contains abundant characteristic information, M E Ayadi et al. [18] divided the phonetic features into four categories, as shown in Fig. 3. When humans use language to express a certain emotion, volume, quality and other characteristics will changed as well. The variety of voice features correlation analysis can better identify the current emotional state. From B Schuller’s systems, we can realize that acoustic, semantic and grammatical features of speech were analyzed and successfully applied to voice interaction in 2002 [19]. W J Han et al. [20] take advantages of ANN to make emotional modelling analysis for the rhythm of the language features and MFCC. Fast correlation-based filter (FCBF) and analysis of variance (ANOVA) are utilized by D Gharavian to extract features for formant, MFCC, energy and pitch in the audio, and then GMM is used to recognize emotion classifications [21]. In addition, voice interaction systems have been widely used in interactive service robots. For example, a humanoid NAO robot has powerful speech recognition abilities, and can realize voice communication with people in the range of two meters smoothly. Its speech recognition system can recognize the contents of the speaker by using semantic analysis [22].

![Figure 3: The classification of speech characteristics](image)

3.3 FEATURE EXTRACTION

Usually there are two main types of emotion Feature Extraction algorithms. One is the emotional analysis method based on single mode state. It applies only to one kind of emotional interaction information modelling and analysis. The other is a multi-modal emotion recognition method. This analysis method is usually combined with different kinds of information, such as: voice and expression, voice and gesture, volume and the semantic, etc. The study of the first method is mature, but it still has such disadvantages as single interactive information, low recognition rate and poor stability. The multi-modal emotion recognition method is closer to human emotion analysis, and it has been applied successfully in several emotional robot interaction techniques. Multi-modal emotion recognition process is shown in fig. 5.

![Figure 5: Multimodal emotion recognition process](image)

In order to achieve better emotional analysis results, many researchers use various kinds of different emotion models to acquire the analysis and judgment of the interaction information. The details are shown in table 1, which are the following cases: (i) The common-used methods detecting and identifying emotion states are not limited to a certain emotion states’ analysis of interactive information, such as the SVM, HMM and ANN; (ii) A lot of sentiment analysis models are based on the typical emotional process improvement and optimization, and even some models fuse a variety of typical models to the analysis of the interactive emotional states.
psychology and reasoning, it makes use of the cognitive human physiological changes detected by using various pu
ning and the man–machine interaction working with human naturally. There are two aspects so far that affective com
ting can be understood. From the physiological point of view, the emotional model is established based on the human physiological changes detected by using various physiological sensors. From the perspective of cognitive psychology and reasoning, it makes use of the cognitive

### 4 Emotional model

The focus of increasing concern in academia is the intelligent robot with human emotions made by affective computing and the man–machine interaction working with human naturally. There are two aspects so far that affective computing can be understood. From the physiological point of view, the emotional model is established based on the human physiological changes detected by using various physiological sensors. From the perspective of cognitive psychology and reasoning, it makes use of the cognitive

#### TABLE 1  The Comparison and analysis of several algorithms of emotion recognition

<table>
<thead>
<tr>
<th>First Author &amp; Year</th>
<th>Input Information</th>
<th>Feature extraction method</th>
<th>Feature recognition method</th>
<th>Database</th>
<th>Emotion analysis model</th>
</tr>
</thead>
<tbody>
<tr>
<td>M Ilbeysi (2012) [16]</td>
<td>Facial expressions</td>
<td>YCbCr color space</td>
<td>Fuzzy Inference System</td>
<td>FACES; MM; Radboud Faces Database(RaFD)</td>
<td>Single mode</td>
</tr>
<tr>
<td>M Srinivasan (2013)[24]</td>
<td>Facial expressions</td>
<td>Discrete Wavelet Transform (DWT)</td>
<td>7-States HMM</td>
<td>A database of 40 face images; FERET</td>
<td>Single mode</td>
</tr>
<tr>
<td>H H Thiago (2013)[12]</td>
<td>Facial expressions</td>
<td>Gabor filter; LBP</td>
<td>SVM, GA</td>
<td>TATFE; Cohn-Kanade</td>
<td>Single mode</td>
</tr>
<tr>
<td>L Chen (2012)[25]</td>
<td>Voice</td>
<td>Fisher Rate; PCA</td>
<td>SVM, ANN</td>
<td>BHUDS</td>
<td>Multi-mode</td>
</tr>
<tr>
<td>W J Han (2009)[20]</td>
<td>Voice</td>
<td>Vector quantization(VQ)</td>
<td>ANN</td>
<td>Own database</td>
<td>Multi-mode</td>
</tr>
<tr>
<td>D Gharavian (2012)[21]</td>
<td>Voice</td>
<td>PCBF; ANOVA</td>
<td>DTW-MLP-GMM model</td>
<td>Farsi</td>
<td>Multi-mode</td>
</tr>
<tr>
<td>Chongben Tao (2013)[26]</td>
<td>Gesture</td>
<td>FNNs; K-means clustering method</td>
<td>MHMMs</td>
<td>Own database</td>
<td>Single mode</td>
</tr>
<tr>
<td>M Rashid (2013)[28]</td>
<td>Facial expressions; Voice</td>
<td>PCA; K-means approach</td>
<td>SVM; Bayes Sum Rule (BSR)</td>
<td>eNERFACE’05</td>
<td>Multi-mode</td>
</tr>
<tr>
<td>Li Zhang (2013)[22]</td>
<td>Facial expressions; Voice</td>
<td>FNNs</td>
<td>ANN</td>
<td>FACS; Cohn-Kanade; ATT-Meta</td>
<td>Multi-mode</td>
</tr>
<tr>
<td>Chao Xu (2013)[29]</td>
<td>Facial expressions</td>
<td>Active Appearance Model(AAM)</td>
<td>HMM; Multi-layer Perceptron(MLP)</td>
<td>Cohn-Kanade; Berlin</td>
<td>Multi-mode</td>
</tr>
</tbody>
</table>

#### TABLE 2  The typical emotional model analysis

<table>
<thead>
<tr>
<th>First Author</th>
<th>Emotional model</th>
<th>Model analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ortony[30]</td>
<td>OCC model</td>
<td>According to the causes of emotions: events, Agent and the object’s emotional tendency, 22 kinds of emotional style are derived by using the universal emotion cognition</td>
</tr>
<tr>
<td>J Velasquez[31]</td>
<td>Catheaxis model</td>
<td>This model contains four subsystems: the behavior systems, emotional systems, motor system and internal stimulation. And it not only considers the relationship between the emotions, but also analyzes the influence of external stimuli on the current emotion.</td>
</tr>
<tr>
<td>Reilly[32]</td>
<td>EM model</td>
<td>A new concept of threshold value is introduced in the emotional model, based on which, to judge the mood is produced by personality characteristics, external stimuli or current state. It has a timely update mechanism of emotional state, which is beneficial to continuous interaction.</td>
</tr>
<tr>
<td>C Breazeal[33]</td>
<td>Kismet model</td>
<td>In this model, the external environment, internal stimulation and actions are judged comprehensively, which will lead to the change of behavior.</td>
</tr>
<tr>
<td>Sloman[34]</td>
<td>CogAff model</td>
<td>By analyzing the information from the reaction layer, transport layer, self-monitoring in the human brain, the model uses emotional processing mechanism to produce the emotional reactions.</td>
</tr>
<tr>
<td>Botelho[35]</td>
<td>Salt &amp; Pepper model</td>
<td>Emotional information is stored by classified and interaction as nodes, using interrupt manager and the generator from cognition, behavior and emotion.</td>
</tr>
<tr>
<td>Miwa[36]</td>
<td>WE-4R model</td>
<td>The emotional system is divided into three layers architecture: reflection, emotion, intelligence, and emotion can be divided into learning system, mood and movement response according to the length of working time. The interplay of reflection and intelligence can get emotional system under the effect of emotion.</td>
</tr>
</tbody>
</table>
In addition, there are some significant emotion models based on Euclidean space, probability space and random event processing theory. Some researchers have made the improvement and optimization based on classic emotional model. For example, W H Kim et al. [37] has made an analysis for the OCC emotion model in a simplified way, and also achieved interaction of the expressions, gestures, and pace on the robot. Embedded emotion model based on the HMM has been put forward by P C Chung [38], and this model can make some analysis of individual behaviour and interactive behaviour through the HMM. Xin Lun [39] has obtained the analysis of emotion model by extending emotion process to continuous space, which enriches the robot’s intermediate state making the man-machine interaction more harmonious.

5 Conclusion

The human-robot interaction technology is a channel for elderly health service robots to communicate with the elderly people. It can make the robot more naturally into the elderly people's life. In this paper, the healthy elderly robot interaction techniques are described and summarized, and emotional models are analyzed in detail. We can see that human-robot interaction technology is gradually developing in the diversity and complexity, and it has achieved phase achievements. But due to the complex process of human emotion analysis, and many factors from external environment, individual features and irrelevant emotion etc. are often needed to consider.

Acknowledgements

This work is supported by National Natural Science Foundation of China (Normal Project No. 61170115), (KeyProject No.61432004) and National Key Technologies R&D Program of China (No. 2014BAF08B04).

References

Jing Han, 1990, Hebei, China
Current position, grades: Ph.D. student at School of Computer and Communication Engineering, University of Science and Technology Beijing.
University studies: University of Science and Technology Beijing

Lun Xie, 1968, Beijing, China
Current position, grades: Professor at University of Science and Technology Beijing and a board member of China Artificial Society, Beijing.
University studies: University of Science and Technology Beijing
Scientific interest: Intelligent robot, Human-robot interaction, intelligent robot and intelligent information processing, embedded system, the automation of process industry.

Shang-mou Xu, 1988, Shandong, China
Current position, grades: master student at University of Science and Technology Beijing, Beijing.
University studies: University of Science and Technology Beijing
Scientific interest: Intelligent robot, Human-robot interaction, Automation of process industry.

Zhi-liang Wang, 1956, Henan, China
Current position, grades: Professor and the head of the electronic engineering department, University of Science and Technology Beijing.
University studies: University of Science and Technology Beijing