IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 22 November 2023, accepted 16 January 2024, date of publication 25 January 2024, date of current version 1 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3358614

RESEARCH ARTICLE

An Interaction Behavior Decision-Making Model of Service Robots for the Disabled Based on Human–Robot Empathy

XINYUE ZHU[1], BOTAO ZHANG[1,2], YUANQI QIU[3], AND SERGEY A. CHEPINSKIY[4]
[1]School of Automation, Hangzhou Dianzi University, Hangzhou 310018, China
[2]Institute of Hydrodynamics and Control Processes, Saint-Petersburg State Marine Technical University, 190121 Saint Petersburg, Russia
[3]HDU-ITMO Cyber-Systems and Robotics Research Center, Hangzhou Dianzi University, Hangzhou 310018, China
[4]Faculty of Control Systems and Robotics, ITMO University, 190000 Saint Petersburg, Russia

Corresponding author: Botao Zhang (billow@hdu.edu.cn)

This work was supported in part by the Fundamental Research Funds for the Provincial Universities of Zhejiang under Grant GK229909299001-004, and in part by the Key Research and Development Project of Zhejiang Province under Grant 2019C04018.

ABSTRACT Currently, most service robots typically receive and execute commands in a passive manner, which is unsuitable for more meaningful Human-Robot Interaction (HRI). In this study, a Human-Robot Empathy Decision-Making Model (HREDM) of service robots is developed for personal assistance services. HREDM contains the perception, cognition, and decision-making that enables the robot to understand the emotions of users and respond appropriately with behaviors that appease or encourage them. First, the SE-ResNet (Squeeze-and-Excitation-Residual Neural Network) is used to recognize and understand users' facial emotions. Then, a Q-Learning-based reinforcement learning model is constructed, which enables the robot to actively learn and interact with users by training on their interaction preferences. The proposed mechanism is used to assess the relationship between the robot's behaviors and the users' emotions and to make decisions to influence the users positively. The experiment results demonstrate that the proposed model allows the robot to actively learn, analyze, and make decisions based on identified emotions, leading to appropriate calming behaviors. Further, it attained a score of 3.7 in a satisfaction assessment with volunteers.

INDEX TERMS Service robots, human–robot interaction, behavior decision-making, emotion regulation, emotion understanding.

I. INTRODUCTION

Service robots are designed to assist and nurse humans [1]. They usually carry out user instructions such as medication delivery and rehabilitation training in hospitals, nursing homes, etc. Unfortunately, most of their interactions lack initiative and naturalness, which can negatively impact the user experience. Humans can understand the emotions of others and can influence them through their behaviors, such as pacifying angry colleagues and inspiring depressed friends. In contrast, most service robots usually only respond to commands given by users. It would be of great significance to the user experience and mental health of the disabled, if service robots could understand the emotions of humans and choose appropriate behaviors to comfort or encourage them, rather than being passive as it is currently. This study aims to establish a human emotion-focused behavioral decision mechanism and utilize a reliable emotion recognition network to understand users for service robots designed for individuals with disabilities. By discovering the correlation between calming behavior and emotions, this mechanism allows robots to empathize with and better support the needs of disabled individuals, rather than solely performing their assigned tasks.

Emotional empathy refers to comprehending and experiencing the other's emotional experience by sharing similar feelings [2], [3]. Most current Human-Robot Empathy (HRE) researches focus on facial expression understanding and how

The associate editor coordinating the review of this manuscript and approving it for publication was Tao Liu.

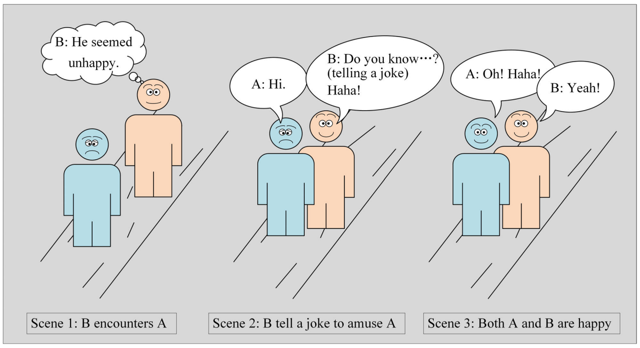


FIGURE 1. **Human empathy. Humans can understand the emotions of others and inspire depressed friends.**

the robot expresses emotions [4]. Meanwhile, psychological researchers suggest a link between empathy and the ability to understand the thoughts of others [5], [6]. This study focuses on finding the correlation between calming behaviors and emotions, which is the role of interactive behaviors in regulating emotions. Emotion regulation involves conscious and unconscious strategies to lessen, maintain, or increase positive or negative emotions [7]. External devices such as robots can be utilized to interact with users to produce emotional stimuli to regulate emotions, in addition to the individual's inhibitions or stimulus responses. And the disabled, a special and large group in society, require services and emotional regulation. Their physical and psychological abnormalities and impairments impact their physical health, interpersonal interactions, and independence, particularly psychologically [8], [9]. Therefore, emotion regulation can be important in maintaining and influencing the mental health of the disabled.

Several studies focused on the importance of emotion and empathy. In chatbots for Human-Computer Interaction (HCI), an emotion-centric approach to emotional response has been proposed to study how to empathize with users in chat responses and then give suitable responses [10]. A multimodal HCI method is used to recognize the user's emotions through expressions, words, and gestures. Then, an agent with designed plots is used to communicate with users through text with the aim of regulating emotions [11]. By combining inference methods and hierarchical analysis, a multi-emotional decision model is constructed, and the algorithm is implemented in a simulation platform for validation and analysis of effectiveness [12]. Additionally, an emotion regulation simulation robot environment is constructed based on a Markov decision process. It weights the analysis of user personality and intent through hierarchical analysis to accomplish the maximum transition from negative to positive emotions while minimizing robot service costs [13]. Researchers have also begun to combine emotional information with other auxiliary information to facilitate behavioral decision-making. A model combining spatiotemporal knowledge of the environment with emotions is constructed, which can reason out the service behavior of the robot, select interaction with users, and verify its effectiveness in a simulation environment [14]. However, the current algorithms' emotion regulation processes are seldom practically applied in real environments, and most of them are experimented with in simulation environments.

Researchers have also attempted to implement the decision-making process by reinforcement learning (RL) algorithms. An approach integrating Q-learning and multi-layer neural networks is employed for coordinated decision-making of multiple-behavior conflicts of robots [15]. A brain computing model inspired by biology is built and integrated with an emotion model based on external rewards and situational memory. Its continuous control of mobile robots is achieved with a model-free approach and a decision method constrained by a global value function [16]. The agent that introduces emotions is proposed in [17] and [18], which uses three emotions, namely happiness, sadness, and fear, as positive and negative enhancement signals to achieve maximum happiness. Nonetheless, they regard the robot as a target for control and ignore the robot's interaction and empathy with humans. A task-oriented dialogue system is built using the Emotion-Sensitive Deep Dynamic Q-learning (ES-DDQ) algorithm. This research developed an emotional modeling approach, designed quick rewards related to emotions, and implemented simulated user behavior [19]. However, it primarily interacts through dialogues and aims to simulate user behavior to produce suitable actions. A reinforcement learning algorithm using fuzzy hierarchical analysis to calculate reward values was applied to perform emotion regulation in HRI in a laboratory and obtained a high degree of satisfaction from experimental users [20]. However, it can only be applied to specific scenarios. The proposed HREDM approach could deal with the above works well in terms of decision-making effectiveness. However, few previous studies refer to empathy and the actual interaction between robots and humans. There is also a lack of consistency in the input and recognition of emotions.

To deal with the above problems, this study attempts to develop a behavioral decision-making mechanism for service robots that is based on the analysis of human emotions. This mechanism enables the robot to choose the proper behaviors to comfort or motivate disabled users according to its understanding of their emotions. Firstly, the Resnet-50 fusion attention mechanism SENet is used to analyze emotions. The results are used as the inputs of the RL model that is based on Q-learning. The RL model is designed for training active interaction, learning user interaction preferences, and analyzing the correlation between the service robot's behavior and the user's emotions. Finally, the decision is made by the model to affect the user's emotions positively. The robot can play an active role in appeasing the user's emotions and changing the blunt and emotionless interaction through this mechanism.

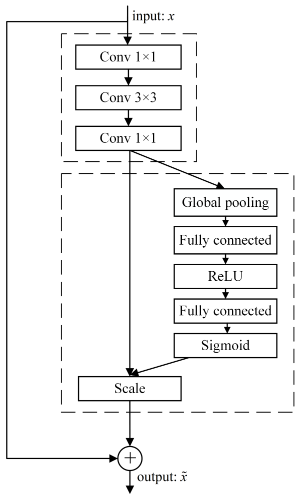


FIGURE 2. The structure of the SE-Residual module. It is composed of a bottleneck followed by an inserted SE-block.

The main contributions of this study are:

(1) The proposed method can enable robots to pay attention to users' emotions and adjust their behavior based on their emotions. By identifying the emotions of users through perceptual and cognitive learning, the proposed model incorporates empathizing with the user and regulating emotions into the interaction and decision-making.

(2) The experimental verification was done in simulation and real-world environments. The feasibility of this approach was evaluated through several applicability experiments, and a near-satisfactory rating was achieved after multiple volunteer experiences.

(3) A facial expression dataset was constructed containing five basic expressions, with a total of 7,829, and utilized for training the emotion recognition network.

II. EMOTION RECOGNITION STRATEGY BASED ON SE-RESNET

This section describes the emotion recognition model, which is mainly based on facial expressions. And the feature extraction capability of the backbone network is enhanced by the attention mechanism module.

Emotions are human attitudes and feelings towards objective things, which are situational, transient, and have more obvious outward expressions [21], and facial expressions are outward expressions of emotions. To achieve accurate facial expression recognition and emotion understanding, it is necessary to extract precise and critical features, because facial expressions are subtle, change frequently, and have slight differences between classes. In this study, ResNet-50 [22] as the backbone network is used for feature extraction, and the SE-block in SENet [23] is introduced to improve the accuracy and efficiency of feature extraction. The SE-Residual module is shown in Fig. 2, and the overall network parameters are shown in Table 1.

TABLE 1. Outline of the SE-ResNet.

Stage	kernel size	output size
input size	-	$64 \times 64 \times 3$
Conv_1	$\begin{bmatrix} 1 \times 1 \times 32 \\ 3 \times 3 \times 32 \\ 1 \times 1 \times 128 \\ fc, [8, 128] \end{bmatrix} \times 3$	$32 \times 32 \times 128$
Conv_2	$\begin{bmatrix} 1 \times 1 \times 64 \\ 3 \times 3 \times 64 \\ 1 \times 1 \times 256 \\ fc, [16, 256] \end{bmatrix} \times 4$	$16 \times 16 \times 256$
Conv_3	$\begin{bmatrix} 1 \times 1 \times 128 \\ 3 \times 3 \times 128 \\ 1 \times 1 \times 512 \\ fc, [32, 512] \end{bmatrix} \times 6$	$8 \times 8 \times 512$
Conv_4	$\begin{bmatrix} 1 \times 1 \times 256 \\ 3 \times 3 \times 256 \\ 1 \times 1 \times 1024 \\ fc, [64, 1024] \end{bmatrix} \times 3$	$4 \times 4 \times 1024$
Global pooling, fully connected	$1 \times 1 \times 2048$	

ResNet with a residual block can solve the overfitting and gradient disappearance problem, which contains residual and short-cut branches. This structure not only enhances the expressiveness of the network and the propagation of gradients but also reduces the memory occupation required during inference. The residual branch of ResNet-50 employs a structure called bottleneck, which is described as follows: first reduces the dimension of the input features after 1×1 convolution, then computes them by 3×3 convolution, and finally uses 1×1 convolution for the dimension reduction to ensure the same dimension as the input. After obtaining the gradient of the residual block output for the loss function $\partial l / \partial y$, the input gradient is

$$\frac{\partial l}{\partial x} = \frac{\partial y}{\partial x}\frac{\partial l}{\partial y} = (\frac{\partial f(x)}{\partial x} + I)\frac{\partial l}{\partial y} = \frac{\partial f(x)}{\partial x}\frac{\partial l}{\partial y} + \frac{\partial l}{\partial y}, \tag{1}$$

where, I is the identity matrix in gradient backpropagation. Connected by a short-cut branch, the output gradient $\partial l / \partial y$ can be passed losslessly to the input $\partial l / \partial x$. This effectively alleviates the phenomenon of gradient disappearance due to the depth of the network, and thus the performance does not deteriorate.

Inspired by the phenomenon that humans selectively devote attention to global information when observing and perceiving things, various networks of attention mechanisms have been proposed in deep learning, with SENet being one such mechanism. SENet enables convolution to focus on the relationships between channels, rather than only local receptive fields. It allows the network to learn the importance of features between different channels. It has the advantage of being highly versatile and can be embedded in most network structures. It first compresses the channel domain, which means that the multiple channels contained in the original feature are compressed into a vector with the dimension of the original number of channels, and the specific implementation process is shown in formula(2).

$$z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} \boldsymbol{u}_c(i, j), \tag{2}$$

where, $\boldsymbol{u}_c$ is the input feature map, H and W are the height and width, respectively, z_c and is the output vector $\boldsymbol{u}_c$ obtained by compression.

Next, the information in the compression operation is used to generate different weights under different channels. The process is shown in formula(3).

$$F_{ex}(z, \boldsymbol{W}) = \sigma(g(z, \boldsymbol{W})) = \sigma(\boldsymbol{W}_2 \delta(\boldsymbol{W}_1 z)), \tag{3}$$

where $F_{ex}(z, \boldsymbol{W})$ is the output after excitation, and $\boldsymbol{W}_1 \in \mathbb{R}^{\frac{C}{r} \times C}$ and $\boldsymbol{W}_2 \in \mathbb{R}^{C \times \frac{C}{r}}$ are the two fully connected layers in the module, respectively, r is the scaling factor, $\sigma(\cdot)$ is the sigmoid function, and δ is the ReLU activation function. After that, the obtained results are used as weights and weighted to the $\boldsymbol{u}_c$ channel by channel to complete the rescaling of the original features in the channel dimension, which is used as the input of the next layer of the network in the form of the formula (4).

$$\tilde{x}_c = F_{scale}(\boldsymbol{u}_c, \boldsymbol{s}_c) = \boldsymbol{s}_c \cdot \boldsymbol{u}_c, \tag{4}$$

where $\tilde{X} = [\tilde{x}_1, \tilde{x}_2, \ldots\ldots, \tilde{x}_c]$, $F_{scale}(\boldsymbol{u}_c, \boldsymbol{s}_c)$ denotes the product of $\boldsymbol{s}_c$ and the feature map.

III. THE ROBOT-ASSISTED SERVICE MODEL BASED ON HREDM

This section describes the decision-making model based on Q-learning and a multilayer robot-assisted service model that includes perception, cognition, decision-making, and execution. The service model is built by combining the decision-making model with the emotion recognition model in Section II.

Unlike deep learning, which relies on the acquisition and construction of datasets, reinforcement learning builds environments for exploring unknown regions and developing training on known regions, eventually arriving at its own decisions to obtain the optimal solution to the problem, as shown in Fig. 3. Among the many reinforcement learning algorithms, the model-free offline learning algorithm Q-learning has attracted the attention of many scholars, and its simple and efficient learning process brings a firm reliance on decision-making [24].

In the Q-learning modeling for decision-making, the agent interacts with the constructed environment and takes different actions from the current state to reach the next state. Different reward values are brought during the transition, and the next state it enters is determined by the current state and reward.

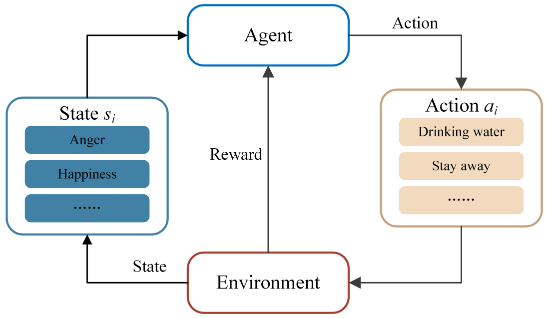


FIGURE 3. Diagram of reinforcement learning based emotion. The robot transitions states by taking actions.

There are four concepts to be illustrated: (1) state space and action space, (2) learning rate, (3) reward, and (4) policy.

State space: The emotional state of the user in the environment. There are five classes of considered emotions. S represents the state space and $S = \{s_1, s_2, s_3, \ldots\ldots\}$.

Action space: An action performed by the agent during the interaction when selecting the next round of appeasing the user's emotional state. The action space contains six classes of actions. A represents the action space and $A = \{a_1, a_2, a_3, \ldots\ldots\}$. The detailed definition of S and A will be elaborated in Section IV.

Learning rate: It can be used to calculate the future reward decay for cumulative rewards of state sequences. It means that the earlier the reward has less impact on the current.

Reward: The reward r can be used to measure the addition of emotion regulation after the agent performs the corresponding action a_i. During the interaction training, different actions of the robot bring different degrees of emotional empathy.

Policy: The policy is used to represent the probability distribution of the agent choosing the next action in the current state. The strategy is used as the action policy as shown in the following Algorithm 1. It means that the action under the maximum is chosen with probability in state s_i, and the other actions are chosen randomly with probability. The purpose of this strategy is to increase the diversity of action choices, explore more possibilities, and ensure that all actions have the probability of being chosen, rather than stopping at the currently considered optimal strategy. Its update strategy uses the greedy strategy. The maximum Q-value in the next state is updated for the current Q-value.

Algorithm 1 ϵ-Greedy Strategy

Input: Random probability ϵ
Output: Action a

1: **if** random $p < \epsilon$ **then**
2: $\quad a \leftarrow \arg\max(Q(s, a))$
3: **else**
4: $\quad a \leftarrow$ random A
5: **end if**
6: **return** a

Algorithm 2 Q-Learning Decision Algorithm Based on ϵ-Greedy Strategy

Input: Training rounds episode *episode*, random probability ϵ, learning rate γ, reward r
Output: Action value function *Q(s,a)*

Initial *Q(s,a)* at random
for $i = 1$ to *episode* **do**
  if s not the terminal **then**
    if random $p < \epsilon$ **then**
      $a \leftarrow \arg\max(Q(s, a))$
    else
      $a \leftarrow$ random A
    end if
    $Q(s, a) \leftarrow Q(s, a) + \gamma[r + \lambda Q_{max}(s, a) - Q(s, a)]$
  else
    $Q(s, a) \leftarrow Q(s, a) + \gamma[r - Q(s, a)]$
  end if
end for
return *Q(s,a)*

Q-learning algorithm is used to implement a service robot interaction behavior decision as shown in Algorithm 2. By establishing a correlation between user emotions and interaction behaviors, the service robot autonomously explores and make decisions about interaction behaviors. It achieves to regulate the user's emotion from negative to positive or maintains positive emotion on the basis of empathy. The algorithm relies on the input emotion as the initial state and selects the subsequent action through the action selection strategy to move to the next state. Then the Q-table is updated using the update strategy, this process is repeated until the training reaches the target epoch and the final decision model, which is the final Q-table, is obtained.

The robot service model is illustrated in Fig. 4. It takes the emotion recognition results inferred by the recognition model as input and interacts with the user using the decision results inferred by the decision model obtained from the training. In the perception layer, the user's facial features are recognized and localized. In the cognitive part, the expression features are extracted and emotion recognition is analyzed by SE-ResNet. The behavior decision results are obtained by Q-learning autonomous learning and training. After that, the decision results are released to the service robot to complete the interaction behavior. Finally, the motion control acts as the execution layer to complete the interaction behaviors through visual positioning, mobile robot navigation, and manipulator and soft claw grasping.

The proposed HREDM is used as the upper layer of the robot service model, i.e. the perception, cognition, and decision-making parts. As an intermediary between users' emotions and interaction behavior, HREDM can effectively empathize with human beings, make appropriate decisions, and finally make the interaction behavior meet the users' preferences. This approach ensures that persons with disabilities have a more positive user experience and emotional engagement in HRI.

TABLE 2. The summary of the emotion dataset.

Categories	Training	Validation	Testing
Neutral	1276	170	168
Happiness	1264	148	165
Anger	1208	169	161
Surprise	1268	176	180
Sadness	1208	128	140

IV. EXPERIMENT

In this section, the experiments are separated into three parts: (1) The emotion recognition experiment based on the self-constructed facial expressions dataset, which is used to verify the dataset typicality and the recognition performance of the model; (2) The applicability evaluation of the proposed HREDM based on (a) a simulation environment and (b) a real service robot for the disabled in a laboratory; (3) The statistics of volunteer test results on HREDM.

A. FACIAL EXPRESSIONS DATASET

The expression dataset is collected based on the laboratory environment and contains five expressions, namely neutral, happiness, anger, surprise, and sadness, which are the basic human expressions that can convey emotions. This dataset collects expression image data from multiple people at different angles, light exposure, and with or without glasses. Because deep learning needs to be trained by extensive data, the data is expanded to fully train the number of network parameters and avoid the overfitting phenomenon. The methods of expansion include randomly flipping, cropping, adjusting the contrast, and adjusting the brightness of the image data during the training input. The dataset is divided into a training set (6224 pictures), a validation set (791 pictures) and a testing set (814 pictures) and summarized in Table 2.

B. EMOTION RECOGNITION EXPERIMENT

Training environment: the operating system is Windows10 64-bit, using the deep learning framework is Tensorflow 1.13.1 based on Python 3.6; GPU is NVIDIA GEFORCE 2070Super (8G), CPU is Intel Core i5-10400. The SE-ResNet network is trained under the above training environment, and the pre-trained model is used for fine-tuning the training, with the number of training epochs is 400, the batch size is 100, and the learning rate set at 0.0001. The accuracy and loss value changes under the test set are shown in Fig. 5. As can

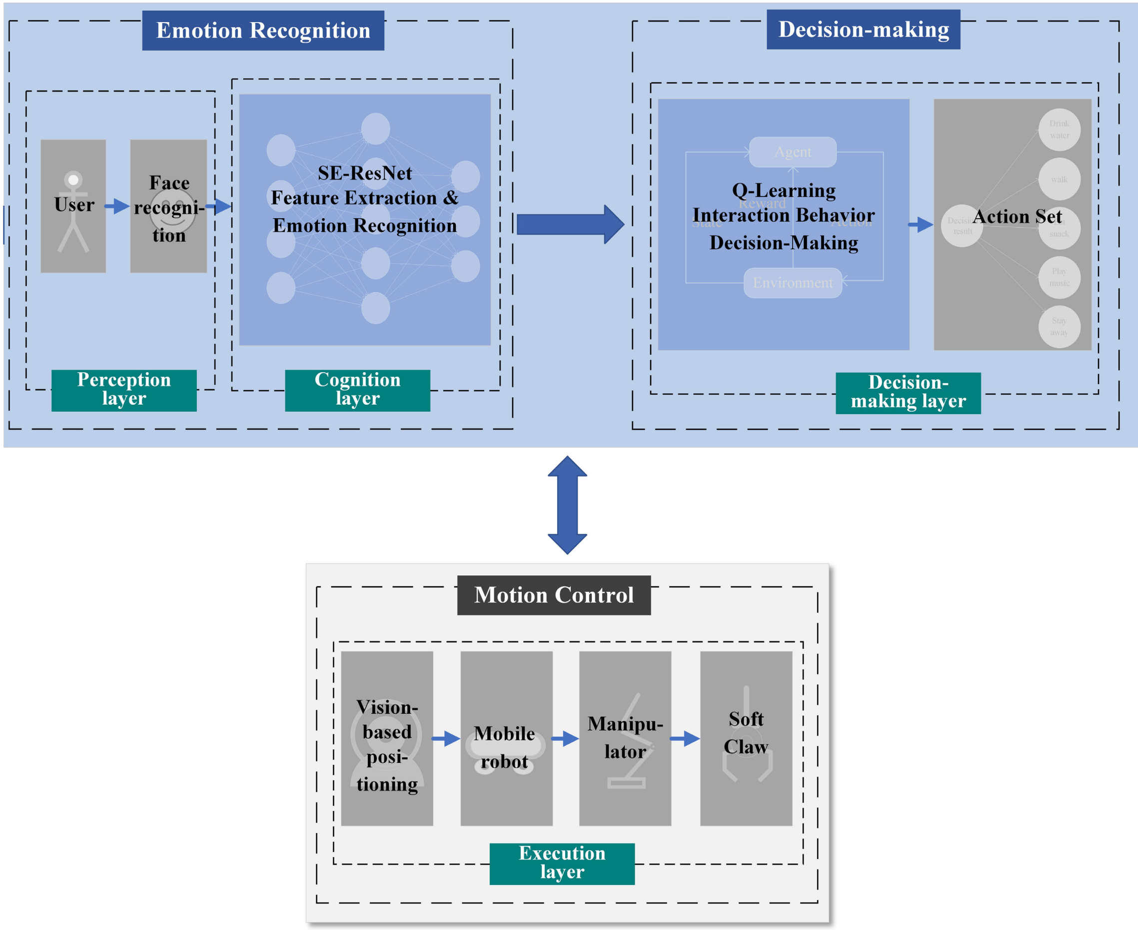


FIGURE 4. The robot service model based on the HREDM. It consists of the proposed HREDM and the execution layer.

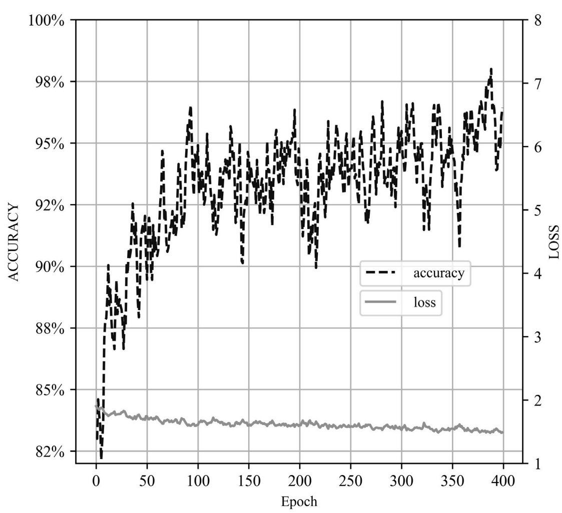


FIGURE 5. The alteration of accuracy and loss on the validation set. The performance of the model is indicated by the stability of the accuracy and the convergence of the loss.

be seen from the figure, because of the pre-trained model, the loss value converges quickly. The recognition accuracy is consistently above 81%, and the final stage of training is maintained between 90% and 98%.

The SE-ResNet used is compared with other current deep learning algorithms as a way to test the recognition effectiveness of SE-ResNet. Since the application mainly relies on the background of disabled assistance, there is no relevant public dataset yet, so the dataset used is self-built. Other comparison methods were used partly with the publicly available datasets CK+ and JAFFEE and partly with their self-built datasets, and the comparative analysis is shown in Table 3. As shown in Table 3, the algorithm used in this paper performs similarly to the current recognition algorithms in terms of accuracy, which is slightly lower than the IF-GAN algorithm by 2.21%. It shows that the algorithm used can recognize users' facial expressions and achieve the goal of emotion recognition and understanding.

C. APPLICABILITY EVALUATION EXPERIMENT BASED ON HREDM

Since interaction training using Q-learning algorithm requires repeated interactions between the user and the robot, the robot can learn the user's interaction habits independently.

TABLE 3. Accuracy comparison of the state-of-the-art methods.

Algorithm	Accuracy
DNN [25]	95.23%
DCNN [26]	95.12%
STRNN [27]	94.89%
CNN+RNN [28]	82.43%
Hybrid CNN-RNN [29]	94.91%
IF-GAN[30]	97.52%
SE-ResNet	**95.31%**

TABLE 4. Definition of Q-learning state space.

Number	State
s_1	Anger
s_2	Happiness
s_3	Neutral
s_4	Sadness
s_5	Surprise

This iterative interaction process can reduce the user's experience and comfort level. Therefore, a simulation environment is designed to implement the training process. Rather than repeatedly expressing a variety of emotions over extended periods, the user can simply input their current emotion within the simulation environment to establish interaction with the robot. The definitions of Q-learning state space is shown are Table 4. Happiness is a positive emotion and is treated as the desired state of training. The corresponding deterministic reward values for the different behaviors are designed as formula (5). The number of training episodes is 100, the random selection probability is 0.9, and the learning rate is 0.01.

$$r = \begin{cases} -1 & s = s_1 \\ +1 & s = s_2 \\ -0.5 & s = s_4 \\ -0.1 & s = else \end{cases} \tag{5}$$

Table 5 shows the action space within two distinct experimental environments. Within the simulated environment, examples of behaviors that are related to the life of the disabled are provided. In the physical environment, behaviors are based on daily needs and experimental conditioning. These behaviors serve as examples to validate the proposed decision-making model. Furthermore, these behavioral spaces can be expanded and extended in accordance with actual scenarios and service objects.

The flowchart of training is shown in Fig. 6, where the purpose of the training is to make the user positive through repeated interactions. At the start of the training, an emotion state will be randomly generated, followed by a choice and the output of a behavior using the ϵ-greedy strategy. A change in emotion will result from this behavior since it will have an emotional impact on the user. When the emotion state changes to "happiness", which is utilized as a positive state objective, the round of training is stopped; otherwise, the interaction training is continued. The generated Q-table will

TABLE 5. Definition of Q-learning action space.

Number	Action	
	expirenment 2(a)	expirenment 2(b)
a_1	eating	staying away temporarily
a_2	cleaning personal hygiene	eating snacks
a_3	rehabilitation	drinking water
a_4	psychological counseling	going out for a walk
a_5	active social interaction	comforting
a_6	physical condition assessment	playing music
a_7	career training	-

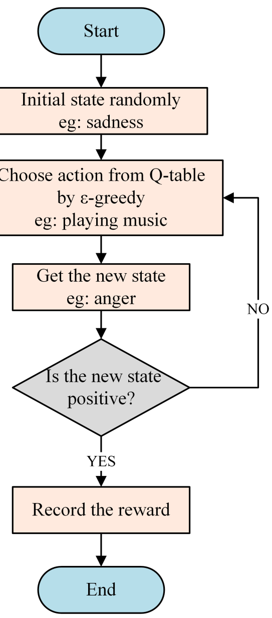
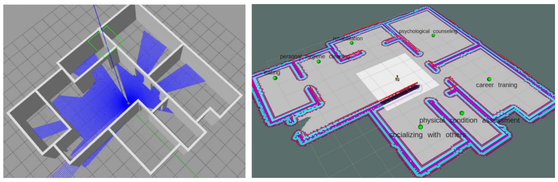


FIGURE 6. Flowchart of training. This represents the process of training within an *episode*. The gray diamond is the *judgment*, if it is a positive state then end the training, otherwise continue to select the behavior.



(a) (b)

FIGURE 7. **Simulation environment. (a) the house built in Gazebo, (b) the corresponding map.**

be able to reflect the relationship between the users' emotions and calming behavior with training. That derived from the training will serve as the foundation for decision-making during the following experimental section.

1) EXPERIMENT BASED ON SIMULATION ENVIRONMENT

A simulated environment was built for testing. First of all, it combines some daily needs of the disabled, which is based on a brief survey of their current situation. There are seven activities of daily needs, including eating, cleaning personal hygiene, rehabilitation, psychological counseling, active social interaction, physical condition assessment, and career training. Then, a house model is built based on the floor plan by Gazebo in ROS (Robot Operating System), and different points in the house correspond to different activities, allowing for clear visualization of decision results. The simulation environment and the map are shown in Fig. 7.

We trained our reinforcement learning model based on these seven activities and obtained the graph of the change in reward values and Q-table as shown in Fig. 8 and Fig. 9, respectively. As can be seen in Fig. 8, the robot initially failed to meet the user's needs. The reward is always negative because the robot has no prior knowledge of the user and selects the interaction behavior at random. The robot learned the user's interaction habits after 50 episodes of training, and the reward change became steady. In Fig. 9, most Q values are negative because all emotions result in lower rewards, except "happiness", which has a positive reward. The Q-table shows that when the user is in a "neutral" state, the robot will consider whether the person wants to "eating", when in a "sadness" state it will ask if it needs "psychological counseling", and when the person is "happiness", it will ask if the user wants to do some social activities with others.

The obtained Q-table is used for the simulation test, during which the camera captures the user's emotion, and then the robot makes a decision based on the Q-table. When the robot receives the result of the decision, it moves to the corresponding position autonomously according to the command signal. A snapshot of the simulation experiment process is revealed in Fig. 10, showing the process when the user is in "anger" and "sadness". The result of the decision "personal hygiene cleaning" is sent to the robot when the user is "anger", and the robot autonomously moves to the point of "personal hygiene cleaning". Meanwhile, the

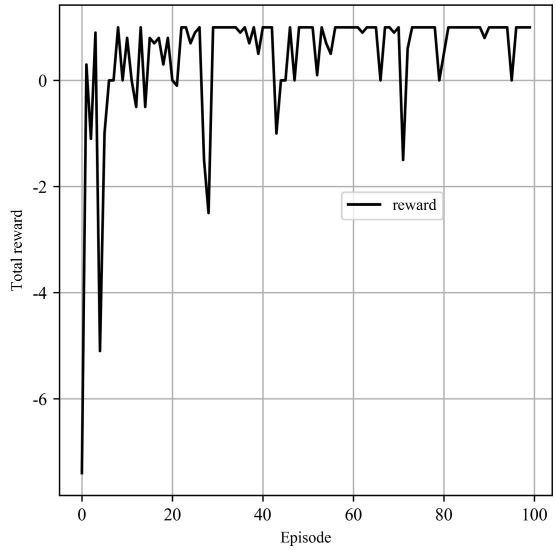


FIGURE 8. **Variation of reward during decision-making training based on the simulation model set-up. The convergence of the model is reflected by the rate of the reward change.**

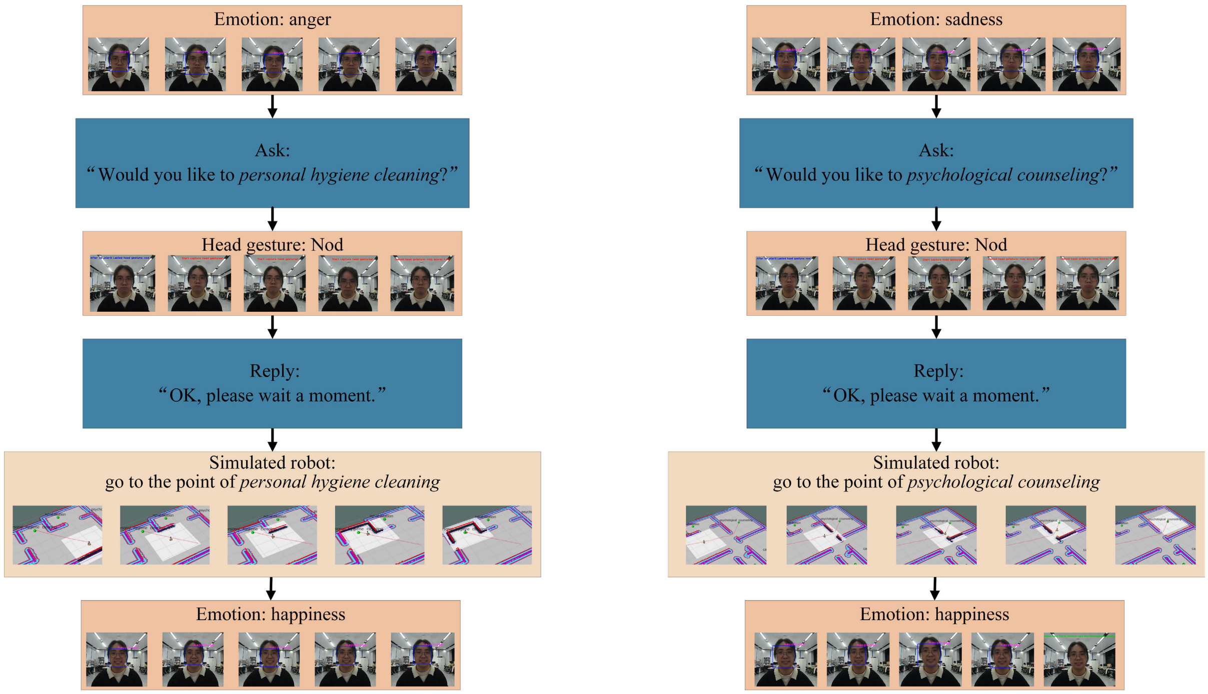
State / Action	Anger	Happiness	Neutral	Sadness	Surprise
eating	-0.010000	-0.001000	0.107024	-0.014985	-0.001999
cleaning personal hygiene	0.152917	-0.009828	0.009363	-0.014938	-0.010997
rehabilitation	-0.010000	-0.014520	-0.001999	-0.018649	0.133231
psychological counseling	-0.010000	-0.009740	0.009010	0.177739	-0.01108
activate social interaction	-0.009617	0.128505	-0.001000	-0.014799	-0.010907
physical condition assessment	-0.018788	-0.010387	-0.001229	-0.014985	-0.002898
career training	-0.020387	0.009662	0.009010	-0.014985	-0.002389

FIGURE 9. **Q-table obtained from decision-making training based on the simulation model set-up. The larger Q(s,a) is, the greater the probability that it will be chosen during the experiment.**

"psychological counseling" is sent when the user is "sadness", and the robot navigates to the point of "psychological counseling". These demonstrate that the proposed model can be utilized in service robots and enables robots to make empathetic decisions based on user's emotions.

2) EXPERIMENT BASED ON REAL ENVIRONMENT

Through training the model on six behaviors in the physical world, we attained a change in reward value during the training process, as illustrated in Fig. 11, as well as the final Q value acquired from the training, as shown in Fig. 12. As can be seen from Fig. 11, the change in the reward value from the initial repeated fluctuation to the final stabilization gradually, indicates that the user's preference tendency has been learned in many training eposides and can make appropriate decisions. In Fig. 12, it is clear that the



(a) (b)

FIGURE 10. Experiment results. Applicability evaluation experiments based on simulation environment when the initial state of the user is (a) anger and (b) sadness.

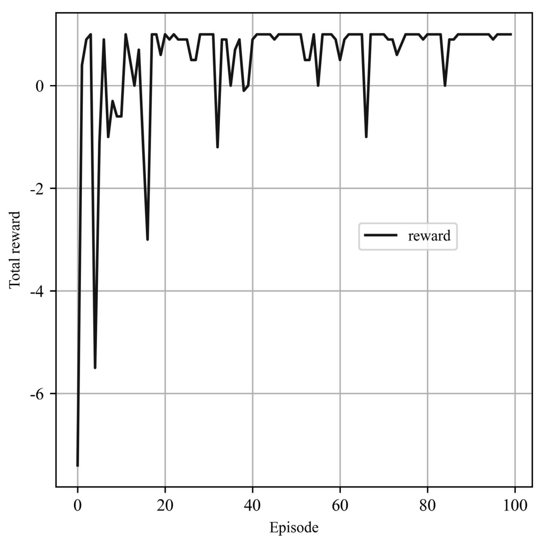


FIGURE 11. Variation of reward during decision-making training based on the real environment set-up.

Action \ State	Anger	Happiness	Neutral	Sadness	Surprise
staying away temporarily	0.114784	-0.010090	-0.011817	-0.004134	-0.011987
eating snacks	-0.029678	0.000979	0.228135	-0.014940	-0.002975
drinking water	-0.019990	-0.010189	-0.001000	-0.014985	0.170904
going out for a walk	-0.019990	0.087821	0.001999	-0.014985	-0.002997
comforting	-0.019791	-0.009839	-0.001999	0.152540	-0.011978
playing music	-0.019990	-0.010312	-0.001347	-0.004835	-0.002997

FIGURE 12. Q table obtained from training the real environment set-up.

users would like the robot to "staying away temporarily" when they are "anger" so that they can calm down alone for a while; when they are "sadness", the robot should "comfort"; and when they are "happiness", the robot's most appropriate behavior is to open the door and let them "going out for a walk". Humans experience the above emotional states along with appropriate psychological expectations, and when expectations are satisfied, the emotional state will change and remain positive. This shows that the model can successfully learn the user's interaction habits and psychological expectations from the interaction process, identify the relationship between emotions and interactions, and change the user's emotions from negative to positive.

The experiment environment, service robot, and interaction system are shown in Fig. 13. We have previously looked into head gesture recognition in [31], allowing persons to communicate their feelings towards a robot through head gestures. The head gestures are introduced into these experiments. Meanwhile, a distributed computing approach is utilized to simulate a service robot that may be remote-controlled within various settings, such as homes and hospitals.

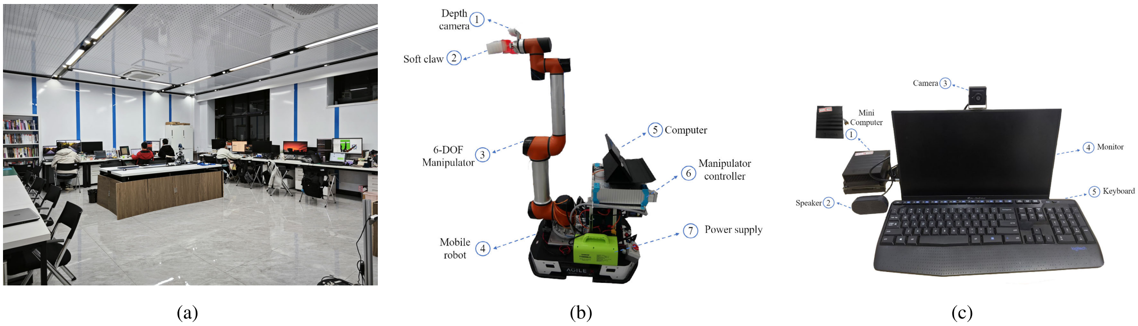


FIGURE 13. Experiment settings. The laboratory environment, service robot, and interaction system.

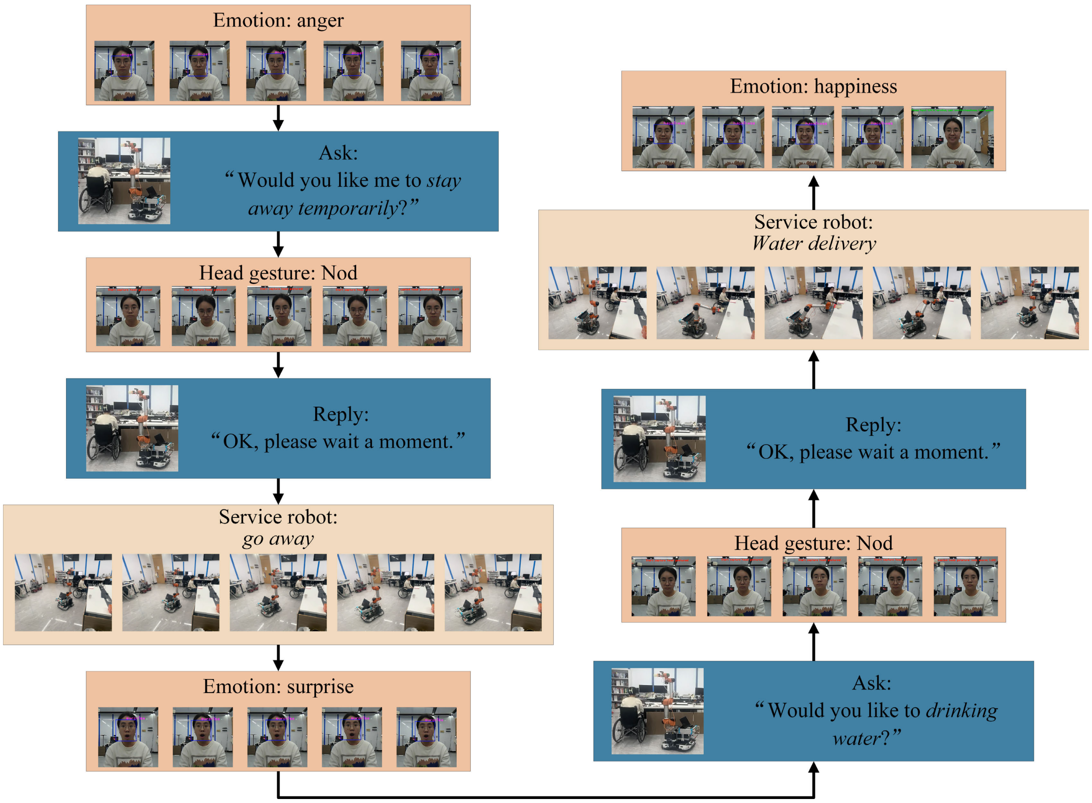


FIGURE 14. Experiment results. Applicability evaluation experiments based on the real environment when the emotion of the user is anger.

Questionnaire on Interaction Behavior Decision-making Experiments

Question 1: Are you satisfied with Interaction Behavior Model 1?
A: Very satisfied (5 points);
B: Satisfied (4 points);
C: Average (3 points);
D: dissatisfied (2 points);
E: Very dissatisfied (1 point).

Question 2: Are you satisfied with Interaction Behavior Model 2?
A: Very satisfied (5 points);
B: Satisfied (4 points);
C: Average (3 points);
D: dissatisfied (2 points);
E: Very dissatisfied (1 point).

FIGURE 15. Feedback questionnaire. Volunteers are asked to rate the satisfaction of both groups of decision models.

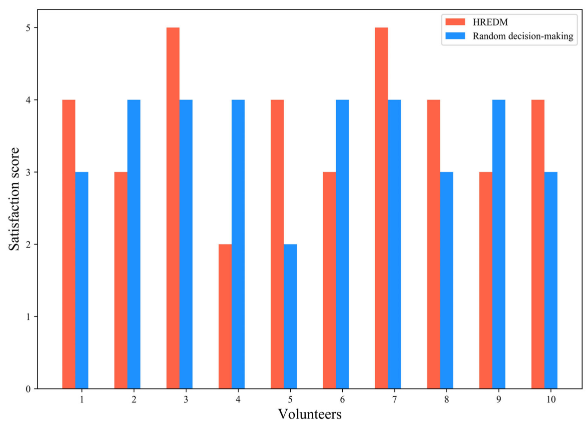


FIGURE 16. Ten volunteers' satisfaction ratings of two interaction models.

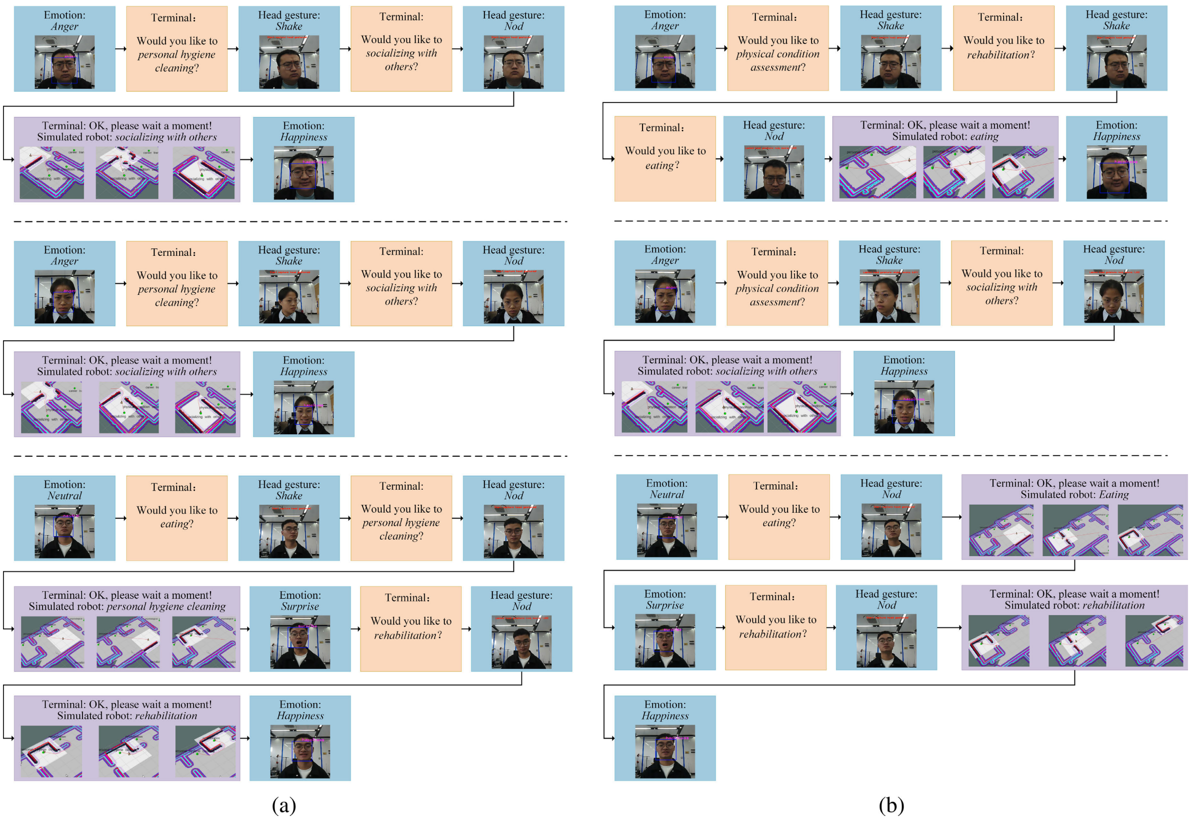


FIGURE 17. Part of the satisfaction experiment snapshot. The interaction process with (a) HREDM and (b) random decision-making. The different users have different interaction tendencies under the same emotion.

The experiment results are presented in Fig. 14. After capturing the user's face and recognizing the emotion, the recognition outcomes are fed into the decision model to obtain empathy results. Results are asked and confirmed to the user by the speaker. After receiving the user's "nodding" response, the current behavior command is published to the mobile manipulator to complete the interaction. The Fig. 14 illustrates the experiment process when the user is angry. Initially, the user's emotion was "anger" and the robot got away from the user by the user's instructions. Subsequently, the user's emotional state transitioned to "surprise". After that, the robot delivers the user with water and tries to bring more positive emotions to users.

The service robot, based on the proposed HREDM, can effectively recognize and responds to the user's emotions and head gestures through its interaction behaviors, which ultimately lead to positive emotion regulation. The results demonstrate that the model can accurately identify the user's emotions and decide the appropriate interaction behavior under the trained model, which brings positive emotion regulation to the user. This model can be applied to a variety of service robots to achieve empathetic behavior decisions based on the user's emotions, appropriately respond to varying emotions, and establish trust and reliance with the user.

D. STATISTICS OF VOLUNTEER TEST RESULTS ON HREDM

Ten volunteers participated in an experiment to the evaluate the validity of the proposed model. The experiment was conducted based on the simulation environment and divided into two groups: empathy-based decision-making(HREDM) and random decision-making. The HREDM group uses the model obtained in section IV, part C(1) of the experiment for interactive decision-making. The group based on random decision-making refers to the random selection of decision-making behaviors in the set of behaviors with an average probability after recognizing the emotion. It means that the decision-making results are independent of emotion. Before the experiment, the volunteers knew that they would need to perform two groups of investigations related to emotion regulation, but they were not told which group of models they were currently in. In this way, the bias of their subjective impressions on evaluations can be reduced. Due to conditional restrictions, the interactive behavior set adopts the behavior set in the simulation experiment part of section IV, part C(1). During the experiment, volunteers can make expressions at will to experience the interaction process. After volunteers have experienced the two interaction models, they need to complete the questionnaire to rate their satisfaction with the two interaction models. The questionnaire is shown in Fig. 15.

The satisfaction score is shown in Fig. 16 and the snapshot of part of the experimental process is shown in Fig. 17. The evaluation results show that the average score of HREDM is 3.7, slightly surpassing that of random decision-making. Two of the volunteers rated the HREDM as ''very satisfied''. These can indicate that the proposed model's capacity for emotional regulation effectively caters to most of the participants. However, one volunteer rated the proposed decision and the random decision as dissatisfied, expressing dissatisfaction because the interaction process took a long time before the desired behavior was decided. It is attributed to the limited range of behaviors in the current training and individual variation among users. In addition, since the comparison algorithms randomly select behaviors with an average probability, the bias of the decided behaviors is significant. Moreover, the willingness tendency of different users under the same emotion varies greatly, so empathy-based decision-making does not score notably superior to random decision-making. It further illustrates that our proposed model can possess an approximate imitation and endeavor of human empathy. Still, space remains for optimization, like augmenting the set of behaviors and including more users to train the model for superior generalizability.

V. CONCLUSION

This study attempted to build a decision mechanism for human-robot interaction called HREDM, which could facilitate more intimate interactions between service robots and the disabled. This mechanism allows robots to comprehend the emotions of their service users, choose appropriate behaviors, and respond reasonably to affect those emotions. In HREDM, the emotion recognition neural network is used to analyze and recognize users' emotions. A Q-learning-based reinforcement learning model was constructed to analyze the correlation between the behaviors of robots and the users' emotions. Based on the above, HREDM makes inferences and decisions to influence the users' emotions positively. The experiment results demonstrate that the proposed mechanism enables the robot to actively find the relationship between interaction behaviors and emotions to please users actively. According to the performance of the proposed approach on multiple volunteers, it indicates that this method can, to some extent, promote users' emotional changes toward more positive aspects. Although it is simply a clumsy imitation of human empathy, this mechanism may provide some insight into conventional HRI. We will investigate HRE and related topics in greater detail.

REFERENCES

[1] J. Wirtz, P. G. Patterson, W. H. Kunz, T. Gruber, V. N. Lu, S. Paluch, and A. Martins, ''Brave new world: Service robots in the frontline,'' *J. Service Manage.*, vol. 29, no. 5, pp. 907–931, Sep. 2018.

[2] A. Smith, ''Cognitive empathy and emotional empathy in human behavior and evolution,'' *Psychol. Rec.*, vol. 56, no. 1, pp. 3–21, Jan. 2006.

[3] M. Asada, ''Development of artificial empathy,'' *Neurosci. Res.*, vol. 90, pp. 41–50, Jan. 2015.

[4] Q. Meng and M. Lee, ''Empathy between human and home service robots,'' in *Proc. IITA Int. Conf. Control, Autom. Syst. Eng. (CASE)*, Zhangjiajie, China, Jul. 2009, pp. 220–224, doi: 10.1109/CASE.2009.156.

[5] M. Iacoboni, ''Imitation, empathy, and mirror neurons,'' *Annu. Rev. Psychol.*, vol. 60, no. 1, pp. 653–670, Jan. 2009.

[6] L. Carr, M. Iacoboni, M.-C. Dubeau, J. C. Mazziotta, and G. L. Lenzi, ''Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas,'' *Proc. Nat. Acad. Sci. USA*, vol. 100, no. 9, pp. 5497–5502, Apr. 2003.

[7] J. J. Gross, ''Emotion regulation in adulthood: Timing is everything,'' *Current Directions Psychol. Sci.*, vol. 10, no. 6, pp. 214–219, Dec. 2001.

[8] C. Tan, W. F. Zeng, X. Zeng, D. Wang, and X. Liu, ''Study on the relationship between the mood state and the cognitive emotion regulation in the disabled,'' *Adv. Psychol.*, vol. 7, pp. 852–857, Jul. 2017.

[9] C. Zhang, W. Guo, and T. Wang, ''A review of the mental health of the disabled in China from the perspective of positive psychology,'' *J. Mod. Special Educ., Res. Higher Educ.*, no. 16, pp. 44–50, Aug. 2018. [Online]. Available: https://kns.cnki.net/kcms2/article/abstract?v=1fhXvtqifPIY5BOU3hdf9nMGklqO-7-MoPfa8dsyE9mVbmyGSuOJYdO4jn1YoeI0Q3PT39wKIPZAKKYDnoRLPJG2qrzMWdDKJkc56q8jISTZBwnEsJss6lljuDjdTuqi2tyWzeaVvyI=&uniplatform=NZKPT&language=CHS

[10] X. Lu, Y. Tian, Y. Zhao, and B. Qin, ''Retrieve, discriminate and rewrite: A simple and effective framework for obtaining affective response in retrieval-based chatbots,'' in *Proc. Findings Assoc. Comput. Linguistics, EMNLP*, Punta Cana, Dominican Republic, 2021, pp. 1956–1968, doi: 10.18653/v1/2021.findings-emnlp.168.

[11] K. Zhang, T. Liu, J. Liu, Y. Zhuang, and Y. Chai, ''Multimodal human–computer interactive technology for emotion regulation,'' *J. Image Graph.*, vol. 25, no. 11, pp. 2451–2464, Nov. 2020.

[12] Y. Shi and C. Li, ''Exploration of computer emotion decision based on artificial intelligence,'' in *Proc. Int. Conf. Virtual Reality Intell. Syst. (ICVRIS)*, Hunan, China, Aug. 2018, pp. 293–295, doi: 10.1109/ICVRIS.2018.00078.

[13] K. Chen, ''Emotion regulation with Markov decision process for human–robot interaction,'' in *Proc. IEEE Conf. Cognit. Comput. Aspects Situation Manage. (CogSIMA)*, Tallinn, Estonia, May 2021, pp. 80–85, doi: 10.1109/CogSIMA51574.2021.9475923.

[14] F. Lu, Y. Jiang, and G. Tian, ''Autonomous cognition and personalized selection of robot services based on emotion-space-time information,'' *Robot*, vol. 40, pp. 448–456, May 2018.

[15] M. Carreras, J. Yuh, and J. Batlle, ''High-Level control of autonomous robots using a behavior-based scheme and reinforcement learning,'' *IFAC Proc. Volumes*, vol. 35, no. 1, pp. 469–474, Aug. 2002.

[16] X. Huang, W. Wu, and H. Qiao, ''Computational modeling of emotion-motivated decisions for continuous control of mobile robots,'' *IEEE Trans. Cogn. Devel. Syst.*, vol. 13, no. 1, pp. 31–44, Mar. 2021, doi: 10.1109/TCDS.2019.2963545.

[17] M. A. Salichs and M. Malfaz, ''Using emotions on autonomous agents. The role of happiness, sadness and fear,'' *Integrative Approaches Mach. Consciousness AISB*, vol. 6, pp. 157–164, Jan. 2006.

[18] M. Malfaz and M. A. Salichs, ''Using emotions for behaviour-selection learning,'' *Frontiers Artif. Intell. Appl.*, vol. 141, pp. 697–698, May 2006.

[19] R. Zhang, Z. Wang, M. Zheng, Y. Zhao, and Z. Huang, ''Emotion-sensitive deep dyna-Q learning for task-completion dialogue policy learning,'' *Neurocomputing*, vol. 459, pp. 122–130, Oct. 2021.

[20] M. Hao, W. Cao, Z. Liu, M. Wu, and Y. Yuan, ''Emotion regulation based on multi-objective weighted reinforcement learning for human-robot interaction,'' in *Proc. 12th Asian Control Conf. (ASCC)*, Kitakyushu, Japan, Jun. 2019, pp. 1402–1406.

[21] W. Xue, ''Research on affective virtual human interaction based on cognitive mechanic,'' *J. Beijing Union Univ. Natural Sci.*, vol. 24, no. 4, pp. 1–6, Dec. 2010.

[22] K. He, X. Zhang, S. Ren, and J. Sun, ''Deep residual learning for image recognition,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778 doi: 10.1109/CVPR.2016.90.

[23] J. Hu, L. Shen, and G. Sun, ''Squeeze-and-excitation networks,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 7132–7141 doi: 10.1109/CVPR.2018.00745.

[24] C. Watkins, ''Learning from delayed rewards,'' Ph.D. dissertation, King's College, Univ. Cambridge, Cambridge, U.K., 1989.

[25] D. K. Jain, P. Shamsolmoali, and P. Sehdev, "Extended deep neural network for facial emotion recognition," *Pattern Recognit. Lett.*, vol. 120, pp. 69–74, Apr. 2019.
[26] W. H. Abdulsalam, R. S. Alhamdani, and M. N. Abdullah, "Facial emotion recognition from videos using deep convolutional neural networks," *Int. J. Mach. Learn. Comput.*, vol. 9, no. 1, pp. 14–19, 2019.
[27] T. Zhang, W. Zheng, Z. Cui, Y. Zong, and Y. Li, "Spatial–temporal recurrent neural network for emotion recognition," *IEEE Trans. Cybern.*, vol. 49, no. 3, pp. 839–847, Mar. 2019, doi: 10.1109/TCYB.2017.2788081.
[28] P. Khorrami, T. Le Paine, K. Brady, C. Dagli, and T. S. Huang, "How deep neural networks can improve emotion recognition on video data," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Phoenix, AZ, USA, Sep. 2016, pp. 619–623, doi: 10.1109/ICIP.2016.7532431.
[29] N. Jain, S. Kumar, A. Kumar, P. Shamsolmoali, and M. Zareapoor, "Hybrid deep neural networks for face emotion recognition," *Pattern Recognit. Lett.*, vol. 115, pp. 101–106, Nov. 2018.
[30] J. Cai, Z. Meng, A. S. Khan, J. O'Reilly, Z. Li, S. Han, and Y. Tong, "Identity-free facial expression recognition using conditional generative adversarial network," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Anchorage, AK, USA, Sep. 2021, pp. 1344–1348 doi: 10.1109/ICIP42928.2021.9506593.
[31] J. Xie, B. Zhang, and Q. Lu, "A dynamic head gesture recognition method based on 3D convolutional two-dream network fusion," *Acta Electonica Sinica*, vol. 49, no. 7, pp. 1363–1369, Jul. 2021.

XINYUE ZHU received the B.S. degree from the School of Automation, Hangzhou Dianzi University, Hangzhou, China, in 2021, where she is currently pursuing the M.S. degree. Her research interests include human–robot interaction and behavior decision-making.

BOTAO ZHANG received the Ph.D. degree from the Institute of Information Science and Engineering, East China University of Science and Technology, Shanghai, China, in 2012. He is currently an Associate Professor with the HDU-ITMO Cyber-Systems and Robotics Research Center and the School of Automation, Hangzhou Dianzi University. His current research interests include robot perception and motion planning.

YUANQI QIU received the B.S. degree from Binzhou University, Binzhou, China, in 2020. He is currently pursuing the M.S. degree with the HDU-ITMO Cyber-Systems and Robotics Research Center, Hangzhou Dianzi University, Hangzhou, China. His research interests include robot control and robot motion planning.

SERGEY A. CHEPINSKIY received the Ph.D. degree in system analysis, control and processing of information in technical systems from ITMO University, Russia, in 2006. Since 2007, he has been an Associate Professor with the School of Computer Technologies and Control, Faculty of Control Systems and Robotics, ITMO University. He is also the Founder of RoboEd Company and the Overseas Expert of the 111 Center, Hangzhou Dianzi University. His current research interests include motion control of mobile robots, robot vision, and underactuated systems.