Toward a Reinforcement Learning Based Framework for Learning Cognitive Empathy in Human-Robot Interactions

Elahe Bagheri[1], Oliver Roesler[2] and Bram Vanderborght[1]

***Abstract*— Observing another's affective state and adjusting one's behavior to respond to it, is the basic functionality of empathy. To enable robots to do this, they need a mechanism to learn how to provide the most appropriate empathic behavior through continuous interaction with humans. To this end, we propose a reinforcement learning based framework for cognitive empathy, which uses reinforcement learning to learn the most appropriate empathic behavior for different emotional states during human-robot interactions.**

To verify the proposed framework, an experiment is conducted with the humanoid robot Pepper over 28 participants, where their facial emotion expression is tracked continuously and used to select appropriate empathic behaviors. The obtained results show the proposed reinforcement learning model converges to the optimal empathic behaviors for all emotions that were expressed a sufficient number of times, which helps the participants feel more positive emotions like happiness.

I. INTRODUCTION

Studies showed robots that can adjust their behavior based on affective state or personality of a user are more accepted as partners for interaction [5] and are seen as more friendly, caring, likeable, supportive and trustworthy [4]. Therefore, different empathic models have been proposed for social robots. However, these models are often special case studies and are evaluated in specific scenarios. For example, Tapus et al. [15] proposed a policy gradient reinforcement learning (PGRL) based model for an assistant robot for a rehabilitation task. The applied reward is the number of exercises performed by the user in the last 15 seconds, which means the model does not take into account the affective state of the user and can only be used in very similar scenarios.

Other studies followed similar approaches by using only contextual parameters of the employed scenarios, like the user's status in playing a game [10] or success in accomplishing a given task, e.g., answering questions [13], to decide whether empathy should be applied or not. However, the current situation of a person is a very indirect and noisy indicator of his/her emotion. Thus, in a previous study [2] we proposed a framework in which the robot extracts the user's affective state from facial expressions and autonomously decides which type of empathy, i.e., parallel or reactive [6], is more adequate for a specific emotional state.

*This work was supported by Flanders make.

[1]Elahe Bagheri and Bram Vanderborght are with Faculty of Applied Sciences, Department of Mechanical Engineering, Vrije Universiteit Brussel, Brussels, Belgium. elahe.bagheri@vub.be, bram.vanderborght@vub.be

[2]Oliver Roesler is with the Artificial Intelligence Lab, Vrije Universiteit Brussel, Brussels, Belgium. oliver@roesler.co.uk

Since empathy is a learnable skill [12] that humans learn through their life by interacting with other humans, an empathic model needs to be continuously updated based on the reactions of different humans. Therefore, empathic models should also incorporate personal preferences of different people regarding the most appropriate empathic behaviors, which might be due to different personalities.

In this study we propose a *scenario-independent learning based empathic model, which learns to select the most appropriate empathic behavior type for different emotional states and types of personality through interaction.*

The remainder of this paper is structured as follows: the proposed framework is described in Section II. The employed experimental scenario and obtained results are discussed in Sections III and IV. Finally, Section V concludes this paper.

II. PROPOSED FRAMEWORK

The proposed framework contains three main modules: (1) Emotion Detection module, which recognizes a user's emotion from facial expressions using the model described in [1], (2) Reinforcement Learning module, which over time learns to select the most appropriate empathic behaviors for different emotional states and personalities, and (3) Empathic Behavior Provider module, which applies selected behaviors to the robot. The two latter components are described in more detail in the following subsections.

A. Reinforcement Learning Module

The learning module, illustrated in Figure 1, uses a "contextual bandit" [14] to enable robots to choose the most appropriate empathic behavior in each possible emotional situation. The number of states equals to the number of considered emotions, i.e., happiness, sadness, anger and surprise, multiplied by the number of considered types of personality, i.e., introvert, ambivert and extrovert (Section II-B), leading to a total of twelve different states. We consider four categories of empathic utterances as possible actions, i.e., Mimical, Motivational, Distractional, and Alleviational utterances (Section II-B), so that the robot learns the most appropriate utterance category for each emotional state.

Based on the situations the users can encounter in the conducted experiment (Section III), we defined "appropriate responses" as responses that change users' emotions from negative to neutral or positive, or let them stay positive and defined the reward accordingly. In this study, happiness and surprise are considered as positive emotions and sadness and anger are defined as negative emotions.

The Q-table is initialized with zeros. At the beginning, the

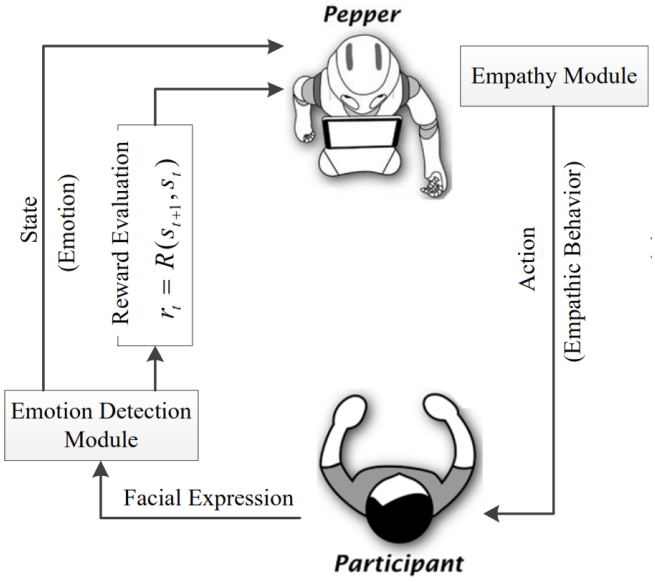

Fig. 1. Illustration of the proposed framework, where the **Participant** (environment) interacts with the robot **Pepper** (agent), the **Emotion Detection Model** recognizes the user's affective state by analyzing **Facial Expressions** (current state), and **Pepper** executes an **Empathic Behavior** (action) selected by the RL model. Afterwards, a reward, which is determined by evaluating the user's new affective state, is given to the applied action.

algorithm starts by selecting random actions, since the Q-values of all utterances are the same. If the new affective state of the user is undesirable, i.e., the user experiences a negative affective state, the Q-value of the selected action decreases so that the action is less likely to be selected again, when the user feels the same affective state in the future. However, if the new affective state is desirable, i.e., the empathic behavior made the user feel positive, the Q-value of the selected action increases so that the robot will choose this utterance in the future with a higher probability. After performing each action, the Q-table is updated based on Equation (1):

$$Q(s,a) \leftarrow Q(s,a) + \alpha[r - Q(s,a)], \tag{1}$$

where a is the action taken in state s, r the corresponding reward, and α the learning rate, which is set to a value of 0.1. For exploration ϵ-greedy is used as described by Sutton et al. [14], where ϵ is set to 0.1.

B. Empathic Behavior Provider Module

The designed empathic behaviors comprise verbal empathic comments, which are categorized as four sets of utterances: *Mimical*, which mimic users' emotions to apply parallel empathy, for instance: "I'm happy that you are happy!". *Motivational*, which are a form of reactive empathy that motivate users to pass the current negative emotion, such as: "Let it go, look for next round". *Distractional*, which can distract users from negative emotions [3], [7], for example: "Do you know what day is today?". *Alleviational*, which try to reduce the user's distress through reactive empathy [9], for instance: "You did your best, don't regret.".

Once the RL model selects the most appropriate utterance category, one of the predefined comments in that category is used randomly. Although empathy depends on the context, the proposed framework is scenario independent since the defined categories, i.e., Mimical, Motivational, Distractional and Alleviational, will be the same in different scenarios/situations, while only the specific comments in each category might need to be adjusted [1].

Moreover, personality can affect the characteristics of verbal comments which are preferred by users, e.g., Tapus et al. [15] showed comments which challenge the user are preferred by extrovert users and comments that focus on nurturing praise are preferred by introvert users. We evaluate users' personality through a questionnaire developed in [8], where the scores in extroversion dimension vary between -5 and 45. We defined three main categories, i.e., introvert, ambivert and extrovert with extrovertion scores under 16, between 16 and 24, and above 24, respectively, because people with small variation in extroversion dimension have similar personalities. Furthermore, as our platform (Section III) is able to express some gestures, we also applied the most related gesture to each comment.

III. EXPERIMENTAL SCENARIO

Although empathy is more appreciated in negative emotional states, receiving empathy in positive emotional states can improve the relation and interaction. In addition, making participants feeling deeply sad is neither ethical nor easy. Therefore, we defined a cooperative interaction scenario, where a robot plays a game with participants. The employed robot is Pepper [11], which is a 1.2 meters tall humanoid robot developed by Softbankrobotics[2]. Pepper has 20 degrees of freedom and is equipped with a tablet, four microphones, touch sensors, LEDs and variety of sensors for multimodal interactions.

In the experiment Pepper asks the participant to tell it which objects it has put in its magic bag in the correct order. Pepper starts the game by saying: "I put in my bag obj_1", where obj_1 is a randomly selected item from a vocabulary set comprises 42 different objects[3], and asks the user to help it to remember what it has in its bag. In response, the participant says "You put in your bag obj_1". We used Google Speech API [16] to track the user's speech and analyse, if the user had repeated all the objects in the correct order. However, due to sensitivity of the Google Speech API to participants' accent, speech speed, and speech loudness, Wizard of Oz confirmed the user's speech in the case the Google Speech API failed in recognizing the user's speech.

Pepper continues the game by adding a new object, and says "I put in my bag, obj_1, obj_2". The game continues until

[1]For instance, at the funeral of a loved one, some replacement comments can be "I am so sorry to hear about your loss." (Mimical), "She had a very nice life and her friends have nice memories to recall her." (Distractional), "We will remember her and keep her memories alive in our hearts." (Alleviational), and "I know that you are strong and I am sure you can handle this tough moment in your life." (Motivational).

[2]https://www.softbankrobotics.com/emea/en/pepper

[3]The vocabulary set contains objects like "book", "tablet", "socks", and "apple".

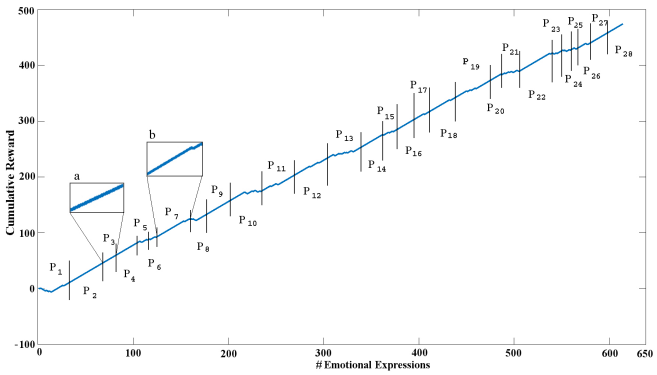

Fig. 2. The obtained cumulative reward by Pepper over all participants, where $P_k(k = 1, ..., 28)$ shows the session corresponding to the k^{th} participant. Box *a* shows the third participant (P_3) liked all of the robot's empathic behaviors since the cumulative reward is ascending, while box *b* shows the seventh participant did not like one of the robot's empathic behaviors as the cumulative reward decreases at one point (the drop in the sub graph).

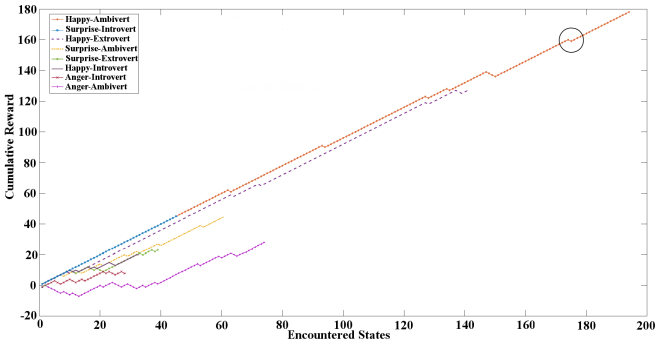

Fig. 3. The cumulative reward that Pepper obtained in different states over the number of times it encountered a specific state. Since different emotions are expressed different number of times, the length of representing lines are different. The circle shows even after the robot converged to the most proper empathic behavior in a specific emotion-personality combination, it is still possible that the robot gets a negative reward once in a while (Section IV).

the vocabulary set gets empty or the participant makes a mistake, i.e., the participant forgets one of the objects or says them in the wrong order. Meanwhile, Pepper tracks the user's facial expressions to detect her emotional state and applies reinforcement learning to select the most appropriate empathic behavior.

The presented results include data acquired in a study with 28 participants, who signed an informed consent form before the experiment. All participants were assigned to one of the defined personality types (introvert=6, extrovert=9, and ambivert=13). Their mean age was 29 years, 19 were male and 9 female. All participants were students or university staff of different disciplines. Participants were compensated with small gifts for their time after the experiment was done.

IV. RESULTS

Since the same learning model is used for all participants, i.e, the model is not trained separately for each individual participant[4], the cumulative reward that the robot obtained by interacting with all the participants, in the order of their presentation in the experiment, is shown in Figure 2. Small black bars are used to separate the different sessions, for instance, since the first participant expressed emotions 33 times, the first black bar intersects the graph where the number of Emotional Expressions equals 33. In this way the reactions of each participant to the robot's empathic behaviors are illustrated, i.e., if the cumulative reward is ascending, it means the corresponding participant liked the robot's empathic behaviors and the robot got positive rewards (box a in Figure 2). If the cumulative reward decreases in one spot, it means the participant did not like one of the robot's empathic behaviors (box b in Figure 2).

As expected, Figure 2 shows that the first participant did not like the robot's empathic behavior, i.e., the cumulative reward is descending, since the Q-values for all of the actions are equal and the RL model selects a random empathic behavior. However, for the second participant, i.e., P_2, the cumulative reward is ascending since the learning model selects the actions that lead to higher Q-values. Figure 2 also illustrates participants' emotion expression behavior, e.g., P_{25} expressed her/his emotions for five times while P_{19} expressed his/her emotions for 47 times. This difference is either because some participants failed in the game so that they had less time to show emotions, or they did not express their emotions so often.

Figure 3 shows the cumulative reward for individual states. For instance, all ambivert participants together got happy for 194 times, while all extrovert participants got happy for 141 times, therefore, in Figure 3, the orange marked line corresponding to Happy-Ambivert goes until Encountered-States equals 194, while the purple dashed line corresponding to Happy-Extrovert goes until the Encountered-States equals 141 since Happy-Extrovert never happened again. The same is the case for other states, i.e., the marked lines representing Anger-Ambivert, Surprise-Ambivert, Surprise-Extrovert, Anger-Introvert, Surprise-Introvert, and Happy-Introvert continue until Encountered-States equals 74, 61, 39, 28, 32 and 45, respectively, which are the number of times each state was encountered.

Due to the employed scenario we did not expect participants to feel negative emotions, which was confirmed by the

TABLE I

THE TYPES OF UTTERANCES WHICH THE MODEL CONVERGED TO IN DIFFERENT STATES ARE INDICATED BY ∗. A, H AND S REFERS TO ANGER, HAPPINESS AND SURPRISE.

Category	Ambivert			Extrovert		Introvert		
	A	H	S	H	S	A	H	S
Mimical		∗	∗	∗			∗	∗
Motivational					∗	∗		
Distractional								
Alleviational	∗							

[4]If the model is trained for each person separately, it cannot generalize across participants, therefore, we only used separate models for the different personalities, i.e., introvert, ambivert and extrovert.

low number of detected negative emotions, i.e., 2 Sad-Ambivert, 7 Sad-Extrovert, 4 Sad-Introvert and 12 Anger-Extrovert. Therefore, the employed RL model could not learn appropriate empathic behaviors for these states, hence, they are not included in Figure 3. In contrast, the learning model converged to the most appropriate empathic behaviors for the other emotional states, i.e., Happy, Surprise and Anger (except Extrovert).

Figure 3 shows that the number of repetitions the model required to converge to the correct empathic behavior for a specific emotion varies. For instance, tracking the blue line representing Anger-Ambivert, the model needed to encounter this state for around forty times to learn the correct empathic behavior, i.e., the cumulative reward remains ascending, while for Anger-Introvert the model seemed to converge after only 18 occurrences of this state, however, it is not clear how stable it is because afterwards the state occurred only ten more times. For positive emotions, i.e., happiness and surprise, the first applied actions got positive reward, therefore, the RL model converged quickly such that it started and continued ascending. Table I summarizes for all of the emotion-personality pairs shown in Figure 3 to which empathic behavior the model converged.

For all states, the model sometimes received negative reward after it had already converged to the optimal empathic behavior, which are shown by drop points in the graphs, e.g., the black circle in Figure 3. This can be the result of one of the following four cases: (a) exploration, which can lead to a non optimal empathic behavior, (b) inaccurate user emotion detection, which can be either that the user's current state is misclassified so that an irrelevant empathic behavior is applied or the user's new state is misclassified so that the received reward is inaccurate, (c) user annoyance due to repetitive empathic behaviors because users might get annoyed, if they hear the same utterance repeated in a short period of time, and (d) the user changed because there is a probability that different participants prefer different empathic behaviors.

V. CONCLUSION

In this study, we proposed a framework to enable robots to learn the most appropriate empathic behavior for different affective states and personalities. The proposed framework determines users' emotional states from facial expressions and uses reinforcement learning to select the most appropriate type of empathic utterance depending on the affective state of the user. To formalize our task as a RL problem, we considered the combination of four emotions and three types of personality as possible states of the environment, defined the type of the provided empathic utterances as possible actions and gave a positive reward, if the user felt positive after the empathic utterance was applied.

Initial results confirmed the ability of the proposed framework in finding the most appropriate type of empathic utterance for considered states. This makes the proposed framework scenario independent, i.e., although the utterances need to be replaced with context related ones, the robot knows which type of utterances are more appropriate for different personality types in different emotional states. However, the number of participants with introvert personality, and also the number of times negative emotions were expressed were not high enough for the proposed RL model to converge in corresponding states. In future work, the experiment will be extended to have more participants, especially introverts, and evoke more negative emotional states to verify the efficiency of the proposed model in this kind of situations. Additionally, we will add speech emotion recognition to the emotion detection module to improve its accuracy of emotion detection. Finally, we will extend the model's state space so that it also considers the user's gender and emotion intensity when selecting empathic behaviors.

ACKNOWLEDGEMENT

The work leading to these results has received funding from the Flanders Make SBO PROUD project and the Flemish Government under the program Onderzoeksprogramma Artificiele Intelligentie (AI) Vlaanderen.

REFERENCES

[1] E. Bagheri, A. Bagheri, P. G. Esteban, and B. Vanderborgth, "A novel model for emotion detection from facial muscles activity," in *Iberian Robotics conference*. Springer, 2019, pp. 237–249.

[2] E. Bagheri, P. G. Esteban, H. L. Cao, A. De Beir, D. Lefeber, and B. Vanderborght, "An autonomous cognitive empathy model responsive to users' facial emotion expressions," *ACM Transactions on Interactive Intelligent Systems (TIIS)*, In Press.

[3] R. P. Bentall, G. Haddock, and P. D. Slade, "Cognitive behavior therapy for persistent auditory hallucinations: From theory to therapy," *Behavior Therapy*, vol. 25, no. 1, pp. 51–66, 1994.

[4] C. N. Brave, Scott and K. Hutchinson, "Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent," *International journal of human-computer studies*, vol. 62, no. 2, pp. 161–178, 2005.

[5] P. Bucci, "Building believable robots: an exploration of how to make simple robots look, move, and feel right," Ph.D. dissertation, University of British Columbia, 2017.

[6] M. H. Davis, "Empathy," in *Handbook of the sociology of emotions*. Springer, 2006, pp. 443–466.

[7] G. B. Diette, N. Lechtzin, E. Haponik, A. Devrotes, and H. R. Rubin, "Distraction therapy with nature sights and sounds reduces pain during flexible bronchoscopy: A complementary approach to routine analgesia," *Chest*, vol. 123, no. 3, pp. 941–948, 2003.

[8] L. R. Goldberg, "The development of markers for the big-five factor structure." *Psychological assessment*, vol. 4, no. 1, p. 26, 1992.

[9] M. L. Hoffman, "Empathy, social cognition, and moral action," *William L. Kurtines/Jack L. Gewirtz: Handbook of moral behavior and development. Bd*, vol. 1, pp. 275–301, 1991.

[10] I. Leite, A. Pereira, S. Mascarenhas, C. Martinho, R. Prada, and A. Paiva, "The influence of empathy in human–robot relations," *International journal of human-computer studies*, vol. 71, no. 3, pp. 250–260, 2013.

[11] A. K. Pandey and R. Gelin, "A mass-produced sociable humanoid robot: pepper: the first machine of its kind," *IEEE Robotics & Automation Magazine*, vol. 25, no. 3, pp. 40–48, 2018.

[12] F. W. Platt and V. F. Keller, "Empathic communication," *Journal of General Internal Medicine*, vol. 9, no. 4, pp. 222–226, 1994.

[13] H. Prendinger and M. Ishizuka, "The empathic companion: A character-based interface that addresses users'affective states," *Applied Artificial Intelligence*, vol. 19, no. 3-4, pp. 267–285, 2005.

[14] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. MIT press Cambridge, 1998.

[15] A. Tapus and M. J. Mataric, "Socially assistive robots: The link between personality, empathy, physiological signals, and task performance." in *AAAI spring symposium: emotion, personality, and social behavior*, 2008, pp. 133–140.

[16] A. Zhang, "Speech recognition (version 3.8)," 2017.