Towards Adaptation of Humanoid Robot Behaviour in Serious Game Scenarios using Reinforcement Learning

Eleonora Zedda [1,2], Marco Manca[2] and Fabio Paternò[2]

[1] *University of Pisa, Largo Bruno Pontecorvo, 3, 56127, Pisa, Italy*
[2] *ISTI-CNR, Via Giuseppe Moruzzi 1, 56127,Pisa, Italy*

Abstract

Repetitive cognitive training can be seen as tedious by older adults and cause participants to drop out. Humanoid robots can be exploited to reduce boredom and the cognitive burden in playing serious games as part of cognitive training. In this paper, an adaptive technique to select the best actions for a robot is proposed to maintain the attention level of elderly users during a serious game. The goal is to create a strategy to adapt the robot's behaviour to stimulate the user to remain attentive through reinforcement learning. Specifically, a learning algorithm (QL) has been applied to obtain the best adaptation strategy for the selection of the robot's actions. The robot's actions consist of a combination of verbal and nonverbal interaction aspects. We have applied this approach to the behaviour of a Pepper robot for which two possible personalities have been defined. Each personality is exhibited by performing specific actions in the various modalities supported. Simulation results indicate learning convergence and seem promising to validate the effectiveness of the obtained strategy. Preliminary test results with three participants suggest that the adaption in the robot is perceived.

Keywords

Social robot, Adaptive Robot Behaviour, Reinforcement learning

1. Introduction

Over the past 20 years, several studies have explored innovative interaction technologies to improve older populations' mental and physical health. From this perspective, there has been increasing interest in robots for usage in social contexts, such as assisting people at work or home with daily activities and healthcare scenarios. For example, social and cognitive stimuli have been found to promote the psychological well-being of older adults and minimise the risk of social isolation, which can negatively impact an elderly individual's health, for example, through increased risk of dementia [1].
Socially assistive robotics [4], which focuses on aiding through social rather than physical interaction between the robot and the user, can improve the quality of life and engagement for large user populations, including the elderly and people with cognitive disabilities [5]. For example, socially assistive robots [2-5] can play the role of conversational companions engaging in conversations with older adults with cognitive impairments. Furthermore, these capacities allow robots to interact more naturally and socially rather than be considered instrumental tools. In this respect, the robot's ability to exhibit personality can be beneficial to support social interaction. Personality represents the set of people's characteristics that account for consistent patterns of feeling, thinking, and behaving [6]. Moreover, different studies [7-8] found that a robot with different personalities can simplify the interaction, as happens in human-human interaction during cognitive training by a human therapist. This is particularly useful when the users are older adults, for example, for performing cognitive training exercises. Indeed, emerging humanoid robots may open up new possibilities in more effectively engaging Mild Cognitive Impairments (MCI) older adults during repetitive cognitive training[5]. In this

Proceedings 2nd Workshop on sociAL roboTs for peRsonalized, continUous and adaptIve aSsistTance (ALTRUIST), December 16, 2022, Florence, Italy.
EMAIL: eleonora.zedda@isti.cnr.it(A. 1); marco.manca@isti.cnr.it (A. 2); fabio.paterno@isti.cnr.it
ORCID: 0000-0002-6541-5667 (A. 1); 0000-0003-1029-9934 (A. 2); 0000-0001-8355-6909 (A. 3)


domain, some contributions indicate that personalized and tailored robotic assistive systems can establish a productive interaction with the user, improving the effects of a therapy session. Various studies used different learning algorithms, including unsupervised, supervised, and reinforcement learning, to develop and design adaptive robot behaviour for supporting older adults during cognitive training. For example, in [9], the researchers designed an adaptive socially assistive robotic (SAR) system that customized a protocol through motivation, encouragement and companionship for users who have Alzheimer's disease. In another study [10], the authors employed Q-learning (QL) to learn a robotic conversation strategy to promote conversation with older adults considering the users' preferred topics and emotions. Adaptive robot systems can be designed and implemented with potential promising results by using various algorithms. In this work, we propose an adaptive robot strategy for adapting the robot's behaviour using Reinforcement learning (RL). RL methods have been successfully applied to model the interaction in problems such as Adaptive Dialogue Systems, Intelligent Tutoring Systems and recently to Robot Assisted Therapy applications [11-12]. The previous studies designed adaptive strategies focusing mainly on exploring robotic dialogue strategies. In our study, we want to find not only the best robotic dialogue but also robot behaviour strategies composed of verbal and non-verbal parameters while exhibiting specific personalities. In this work, we focus on designing an adaptive strategy to maintain the user engaged while the robot performs a personality according to the user state detected using Q-learning (a model-free reinforcement learning algorithm).

2. Adaptive Robot Application

In this section, we describe how we have applied an RL technique to support an adaptation strategy for the robot in the context of an application for cognitive training. The goal is to help the user to maintain a positive attentive level and stimulate in case the user is at a low attentive level.

2.1. Aspects of the Pepper Robot behaviour considered for the study

The humanoid robot used in this work is the Pepper, developed by Softbank's Robotics. Pepper is a 1.2-m tall humanoid robot with 17 joints for expressive body language and three omnidirectional wheels to move around. Pepper has multimodal interfaces for interaction: touchscreen, speech, tactile head, hands, bumper, LEDs and 20 degrees of freedom for motion in the whole body. The robot is equipped with an LG CNS screen of 10.1 inches with a resolution of 1280x800 for supporting touch interaction. Pepper is equipped with motors that allow it to move the head, arms and back, six laser sensors and two sonars, which allow it to estimate the distance to obstacles in its environment. The robot can detect various user states using a combination of sensors. We used the user gaze direction and the smiling state in this work. For the gaze direction values, we consider: (1) looking at the robot, (2) looking up, (3) looking at the tablet, (4) looking left, (5) looking right. The user is in the "attention state" when is looking at the robot or at the tablet (whereas he/she considered in a "distracted state" for the other gaze directions). The smile state is categorized into three values: (1) not smiling, (2) smiling and (3) broadly smiling. The defined parameters will be collected during the interaction with the robot using the modules offered by the QiSDK robot framework. In order to obtain enough data to assess the cognitive state, data will be collected every 2 seconds.

We represent a user's current state by combining these values with the answers given by the user while interacting with a serious game. In this scenario, the user state may change depending on the robot's actions. We considered an action a combination of verbal and non-verbal parameters per the robot's personality. In our work, the robot exploits two different personalities: an extraverted personality and an introverted personality. Studying the state of the art, we extrapolate different parameters that allow the robot to manifest such two opposite personalities [13]. Usually, extroverts tend to speak louder, faster and with a higher pitch. Typically, they are more inclined to initiate conversations and speak more about themselves than others. Regarding gestures and movements, they are usually wider and faster and occur more often than those of an introverted person. We consider in this scenario three possible actions. If the user is attentive and gives the correct answer to the question asked by the robot, the robot can exhibit appropriate, engaging and enthusiastic behaviour. While if the user is not attentive but provides a correct answer, the robot should provide behaviour that does not lead the user into a less

attentive state. On the other hand, if the user is in a "bad" moment, the robot should provide stimulating behaviour to encourage the user to stay attentive and try to re-engage the user.

2.2. Scenario application

The adaptation will guide the robot in which action choose according to the user state detected during a serious game. The cooking game consists of 8 questions requiring users to recognize the ingredients' sequence and the weight's ingredients. The serious game can be into five states: introduction, recipe instruction, question state, answer state, and ending feedback. When the application starts, the robot greets the user and asks if it is ready to play. When the cooking game starts, the robot shows and vocally synthesizes the ingredients for the selected recipe. The robot emphasises the sequential ingredients' order and weight during the recipe instruction. Then, it starts the quizzes, during which the user should use visual attention and working memory to recognise the right ingredients and select them among other options available. Finally, the user has to guess the answer among the four elements proposed. The user interacts through the voice modality. The user state is collected and evaluated for each of the eight questions, and according to that specific value, the robot will perform the optimal action for that state based on the indications of the RL algorithm. We decided to use it in this scenario because the robot adaptation is a crucial element to engage the user more during the serious game. An engaged robot adaptation is important in this context because, typically, the user is exposed to a series of repeated and standardised tasks with challenges that target specific cognitive domains and may create a high risk of dropping out of therapy and the generation of adverse conditions in the users, particularly older adults.

2.3. Definition of the reinforcement learning support

As shown in Figure 1, interactions within the RL setup are sequences of states *S*, actions *A* and rewards *R*. More specifically, the RL agent perceives state S_t from its environment, based on which it selects to execute action A_t. The action is executed, and the environment returns a new state, S_{t+1}, with a reward, R_{t+1}, which evaluates the current transition. The agent selects actions based on its policy π, which maps states to actions and dictates which action to execute given the current state. The goal of the agent is to interact with its environment by selecting actions in a way that maximizes future rewards.
In our work, we used Q-learning. It is a model-free, off-policy reinforcement learning that aims to find the best course of action given the agent's current state. Depending on where the agent is in the environment, it will decide the next action to be taken.
In Q-Learning, the policy is expressed as a *Q table(s, a)* matrix, where *s* is the environment's current state, and *a* denotes the robot's action choice to interact with the user. All actions are defined in the robot's action space *A*. In our case, an action ~ $a \in A$ can be a combination of verbal feedback, vocal parameters, animations and motor movement that the robot provides for personality. In our case, the state space *S* can be any value that describes the user's state, which is composed of the attention level, identified through the user gaze direction and smile state, and in addition, the rightness or wrongness of the user answer given after each question asked by the robot.

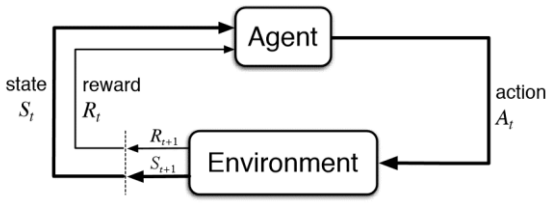

Figure 1 R-learning process

According to our goal for a more engaging robot behaviour generation, we define the key elements of the Q-learning algorithm as follows:

2.3.1. State space

A state is defined according to a user's situation during a serious game scenario. The user state is defined as the combination of five values for the user gaze, three values for the user smile state and the answer given by the user (right or wrong). Based on the designed model of the user state, the state space has a cardinality of 5x3x2=30.

2.3.2. Actions

The Actions correspond to the robot's behaviour (verbal feedback, vocal parameters, animation and motor movements) per the robot personalities supported. During the serious game application, the robot may take three actions. (a_0) Generation of a more engaged and energetic behaviour. For example, in the extravert personality, the robot will provide more enthusiastic feedback such as: "My gosh! That is the correct one! You are trying hard!" slightly increasing the speech speed, volume, and pitch; with a more dynamic and extensive animation with more significant motor movements. (a_1) Generation of more neutral robot behaviour. An example is: "Good! That is the right answer! "with neutral vocal parameters defined by its personality and with a basic animation according to the robot's personality performed. (a_2) Generation of a more stimulating behaviour. An example, "right answer! Let us continue with this focus!" and more close animations. Table 1 shows examples of the parameters used for each action in the extrovert personality.

Table 1

Example actions for extravert personality

	Action 0	Action 1	Action2
Vocal feedback	Is wrong! Come on; we can do it!	Is wrong, try again!	Is wrong! Come on! Maximum concentration! Let us frame the answers better!
Vocal parameters	Vocal speed + 10%, pitch + 5%	Vocal speed and pitch set as personality default	Vocal speed -2%, pitch -2%
Animation	Broadly animations with big angles	Animation defined for that personality	Slightly more closed and stationary animations
Motor Movements	Movement forward and diagonal ~18 cm	Neutral movement. Forward movement ~13 cm	forward and diagonal movements ~5 cm

2.3.3. Reward

According to the goals for a social robot in the cognitive game scenario, i.e. stimulate the user to maintain a high level of engagement (evaluated with the user gaze and the smile state) and to stimulate the user to focus on the questions, we have built our immediate reward function as in Table 2. The reward function considers the user's answer given for each question, the user's gaze direction and the user's smile state. Specifically, the robot should always try to prevent the user from getting trapped in a low level of attention (not looking at the robot and not smiling) and getting the wrong answer.

Accordingly, the reward component of the user's gaze direction, looking at the robot, is set to +5, while if the user is not looking at the robot, the reward is set to -5. The reward component of the smile condition is set to –1 for not smiling, +1 for smiling and +5 for broadly smiling. The reward wants to give more weight to the user's gaze direction than the smiling state because we consider it more important when the user looks for the estimation of an attentive state. The reward component for the answer has a different weight based on the robot's action. This is because if the user makes a mistake in an attentive state, the decline to a non-attentive state is higher (R= -15). In comparison, the rise from an inattentive to an attentive state is slightly slower (R=+8) because the users have to demonstrate being attentive, and more than a correct response is needed to get them back to action A0.

Table 2

Reward values for user state

Value	User State-Gaze	Reward

[1]	Look at the robot	+5
[3]	Look at the tablet	+5
[2,4,5]	Look up, left, right	-5

Value	User State-Smile	Reward
[1]	Not smile	-1
[2]	Smile	+1
[3]	Broadly Smile	+5

Value	User State	Reward A_0	Reward A_1	Reward A_2
[1]	Right	+15	+10	+8
[0]	Wrong	-15	-10	-8

2.4. Preliminary Experiment

In this section, we present our preliminary adaptation experiment to obtain the optimal q-table that the robot use to choose which action is the more suitable for that user state and the results of the interaction of three users with the robot behaviour adaptation. In particular, this adaptation aims to maintain a high level of attention in the user and stimulate the user more if he falls into a negative "mood". As we mentioned, we assume that user attention level relates to the different robot behaviour adaptations; if the robot selects the appropriate action, the attention should be high. Based on the definition above for the three key elements (state, action, and reward function) of Q learning, we have trained our reinforcement learning model in python. The RL agent has been trained for 1500 epochs, each with eight episodes. The learning rate and discount factor have been set to 0.5 and 0.2, respectively. We used at the begging an exponential ε decay; then we set the ε-greedy policy at ε =0.2. An episode starts when the robot asks the first question and ends when the user answers the last question. We evaluate the performance of our model with the sum of Q-value updates during each epoch, and the Q table mean over the number of epochs to evaluate the convergence of performance (Figure 2). From Figure 2, we observe that the algorithms reach the convergence to the optimal policy after 400 epochs. Thus, reaching the convergence, additional training will not improve the model.

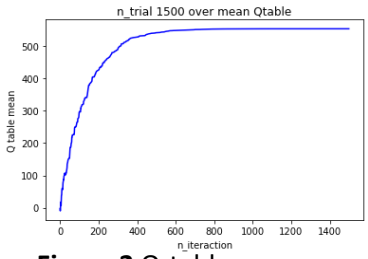

Figure 2 Q-table mean

In a within-study design, the proposed adaptation robot behaviour was tested with three users (2 males) between 28 and 45 years old (M = 36.33, SD= 6.09). The requirements for enrolling were to be at least 18 years old and Italian-speaking. For the test, the users interacted one by one in the lab, sitting in front of the robot. The experiment took an average of 45' for each user (test + semi-structured interview). A moderator took notes of user feedback and user behaviour during the test. All users were exposed to both conditions (interacting with robot adaptations and without adaptation performing the extravert personality). The robot randomly chooses one of the three actions to perform with no adaptation condition. The semi-structured interview was composed of four questions regarding whether users had perceived differences between the two types of robot behaviour, the likeability of the two types of robot behaviour and the positive and negative aspects of both interactions. As a result, we obtained that all three users have seen the adaptation in the robot. The users highlighted main elements for the adaptation are the pitch change, the speech rate and more engaging and stimulating feedback. They described the adaptive robot as more stimulating and active, while the nonadaptive robot was considered extroverted but more neutral and "standard". The positive aspects highlighted for the adaptive were that it seemed very personalized, and they liked the changes in voice dynamics and movements. The nonadaptive robot was also liked for its calmness in the robot's motor movements. A

negative aspect of the adaptive robot was the overly pronounced forward movement that was also seen in the nonadaptive robot but was less annoying in that case.

3. Discussion and Future Work

In this work, we have designed and defined the parameters for supporting an intelligent algorithm to manipulate the robot's behaviour. We identify three possible actions that can be used to stimulate and increase user attention during a serious game scenario. For this reason, we have trained a Q table used in the intelligent algorithm to drive the robot by identifying the best action to perform based on the detected user state. In training the algorithm, we simulated an older adult's interaction, deciding which states to reward and which to penalize to bring the user to a positive state. We performed a preliminary experiment to evaluate the perception of the robot adaptation with real users. From this preliminary study, the participants perceived differences between the two robot behaviour and identified which robot was more adaptive and stimulating their attention. Future work will be dedicated to empirically validating the robot adaptation in a user test with a larger sample of users and with older adults.

4. References

[1] Silva TBL, Dos Santos G, Moreira APB, Ishibashi GA, Verga CER, de Moraes LC, Lessa PP. Cognitive interventions in mature and older adults, benefits for psychological well-being and quality of life: a systematic review study. Dement Neuropsychol. 2021 Oct-Dec;15(4):428-439.

[2] Carros, F., Meurer, J., Loffer, D., & Unbehaun, D. (2020). Exploring human-robot interaction with the elderly: results from a ten-week case study in a care home. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, (pp. 1-12).

[3] Pino, O., Palestra, G., Trevino, R., & De Carolis, D. (2020). The humanoid robot nao as trainer in a memory program for elderly people with mild cognitive impairment. International Journal of Social Robotics, 12(1), 21-33.

[4] Beuscher LM, Fan J, Sarkar N, Dietrich MS, Newhouse PA, Miller KF, Mion LC. Socially Assistive Robots: Measuring Older Adults' Perceptions. J Gerontol Nurs. 2017 Dec 1;43(12):35-43. doi: 10.3928/00989134-20170707-04.

[5] Manca, M., Paternò, F., Santoro, C., Zedda, E., Braschi, C., R., F., & Sale, A. (2021). The impact of serious games with humanoid robots on mild cognitive impairment older adults. International Journal of Human-Computer Studies, 145, 102509

[6] Pervin, L. A., Cervone, D., John, O. P. (2005). Personality: Theory and Research (9th ed.). Hoboken, NJ: John Wiley and Sons

[7] Adriana Tapus, Cristian Ţăpuş, and Maja J Matarić. 2008. User—robot personality matching and assistive robot behavior adaptation for post-stroke rehabilitation therapy. Intelligent Service Robotics 1, 2 (2008), 169–183.

[8] Sarah Woods, Kerstin Dautenhahn, Christina Kaouri, René te Boekhorst, Kheng Lee Koay, and Michael L Walters. 2007. Are robots like people?: Relationships between participant and robot personality traits in human–robot interaction studies. Interaction Studies 8, 2 (2007), 281–305

[9] Tapus, A.: Improving the quality of life of people with dementia through the use of socially assistive robots. In: Advanced Technologies for Enhanced Quality of Life (AT-EQUAL 2009), pp. 81–86. IEEE (2009)

[10] Magyar, J., Kobayashi, M., Nishio, S., Sinˇc´ak, P., Ishiguro, H.: Autonomous robotic dialogue system with reinforcement learning for elderlies with dementia. In: 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC). pp. 3416–3421. IEEE (2019)

[11] Chi, M., VanLehn, K., Litman, D., Jordan, P.: An evaluation of pedagogical tutorial tactics for a natural language tutoring system: a reinforcement learning approach. Int. J. Artif. Intell. Educ. 21

[12] Modares, H., Ranatunga, I., Lewis, F.L., Popa, D.O.: Optimized assistive humanrobot interaction using reinforcement learning. IEEE Trans. Cybern. 46, 655–667 (2015)

[13] Zedda, E., Manca, M., & Paternò, F. (2021). A Cooking Game for Cognitive Training of Older Adults Interacting with a Humanoid Robot. In *CHIRA* (pp. 271-282).