

A Proposal for Adapting Robot Behaviours Using Fuzzy Q-learning in Cognitive Serious Game Scenarios

Eleonora Zedda^{1,*}, Fabio Paternò¹

¹*HIIS Laboratory-Institute of Information Science and Technologies "Alessandro Faedo" (ISTI-CNR), Via Giuseppe Moruzzi, 1, 56127 Pisa PI, Italy*

Abstract

The repetitive and monotonous character of cognitive training may lead to waning interest and eventual disengagement among older adults with cognitive impairments. To address this issue, this study proposes an adaptive approach wherein a Socially Assistive Robot (SAR) autonomously selects optimal actions to sustain an emotional state in older adults while participating in serious cognitive training games. The aim is to propose an adaptation strategy that leverages fuzzy Q-learning to prompt users to maintain a positive state.

Keywords

Human-Robot Interaction, Robot behaviour adaptation, Socially Assistive Robots, Fuzzy Q-Learning,

1. Introduction

Over the past two decades, numerous research efforts have delved into innovative interaction technologies to improve older individuals' mental and physical well-being. In this context, there has been a growing interest towards integrating robots into social environments, intended to help individuals in both professional and domestic contexts, specifically in cognitive training. In particular, it has been observed that exposure to social and cognitive stimuli can significantly strengthen the psychological well-being of the elderly and, at the same time, mitigate the dangers of social isolation, a phenomenon with harmful effects on the health of the elderly, which can lead to a 'high susceptibility to conditions such as dementia[1]. To facilitate natural interaction, researchers in social robotics have focused on robots that can adapt to diverse conditions and different user needs [2]. Machine learning techniques for adaptable social robots have recently attracted a lot of attention [3] [4] [5] [6] [7]. Adaptive robot interactions are essential to provide comfortable, effective, and affordable interactions with humans. An adaptive behaviour system would facilitate meaningful, effective communication and interaction. Additionally, it can create a more trusting relationship between the user and the robots [6, 8]. To enable machines to interact with users naturally, the system must be able to identify or recognize the state of human behaviour and performance through the input modalities (e.g., speech recognition, emotion detection, gestures, etc.) and com-

municate through its output modalities (e.g., speech and gesture generation). Previous research has predominantly focused on devising adaptive strategies for exploring robotic dialogue techniques and robot communication atmosphere using Reinforcement Learning (RL) or using Fuzzy Q-learning [2, 8, 9]. Our investigation aims to identify the appropriate robot behavioural strategies, encompassing verbal and non-verbal parameters, by infusing specific personalities into a socially assistive robot (SAR) interacting with older adults suffering from mild cognitive impairment (MCI). The rationale for this emphasis lies in existing literature [10, 11, 12], which suggests that customizing and adapting the SAR system can produce more effective and engaging human-robot interaction (HRI), especially within a vulnerable demographic such as older adults with MCI. The proposed strategy combines the potential of Q-learning (QL) and that of Fuzzy Logic, which allows the management of a fuzzy number of states. In this proposal, the user state is not limited to a small and discrete set of states in the serious game scenario, as in RL. With the potentiality of fuzzy logic, the set of states is more generalized and fuzzy to reflect a user's natural and nuanced state during the interaction with the robot. In addition, fuzzy logic is well known for its successful application to uncertain environments[9]. Thus, providing adaptive behaviour to the robot is crucial for this particular population, as highlighted by previous studies, in which both older adults individuals and caregivers expressed a preference for more natural and responsive interaction with the robot during cognitive training sessions [1, 13, 14].

Workshop Robots for Humans 2024, Advanced Visual Interfaces, Arenzano, June 3rd, 2024

*Corresponding author.

✉ eleonora.zedda@isti.cnr.it (E. Zedda); fabio.paterno@isti.cnr.it (F. Paternò)

📄 0000-0002-6541-5667 (E. Zedda); 0000-0001-8355-6909

(F. Paternò)

© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2. Approach and Motivation

Reinforcement Learning is a fundamental learning paradigm within machine learning[15]. It offers a stan-

standardized framework for developing agents capable of acquiring optimal behaviour within uncertain environments. Within RL, agents operate without direct access to an optimal control strategy, relying instead on instantaneous reinforcement signals. Most RL algorithms [16] commonly depict the action-value function using a look-up table, allocating a singular entry for each state-action pairing. While this method boasts robust theoretical underpinnings [15, 17] and proves effective across various applications, its utility is severely constrained when confronted with problems featuring extensive state spaces or continuous domains, owing to the phenomenon referred to as the curse of dimensionality [16]. To address this challenge, numerous methodologies attempt to mitigate such limitations by employing function approximation techniques such as Fuzzy Logic, thereby approximating the action-value function using limited parameters [18, 19]. Consequently, the agent's exposure is confined to a set of the state space, whereby, through the generalisation mechanism, it can yield a satisfactory approximation across a broader expanse of the state space. To address such challenges, a promising approach is to model the adaptation simulation using Fuzzy Q-learning (FQL). FQL is an extension of the original Q-learning proposed in Reference [17] that uses fuzzy parameters [20, 21]. Introducing Fuzzy Q-learning allows us to consider more prominent and variable ranges of values than the choice of a specific value. This adds more variability and representation in the simulated users. We propose a Fuzzy RL technique to support an adaptation strategy for the robot, composed of verbal and nonverbal parameters while exhibiting specific personalities, in the context of an application for cognitive training [22].

2.1. Key Element in the Proposal

The adaptive strategy is applied to a cognitive training scenario in a serious cooking game. The robot manifests an extraverted or introverted personality during the interaction between the user and the cooking game. The choice of this personality, made before the training session, modifies the robot's behaviours following pre-established parameters. The adaptation algorithm determines which actions, associated with the current personality, will be selected based on the maximum value that maps the user's current state and the action. In the following scenario, there are three potential actions that the robot can take. Suppose the user is in an extremely positive state (e.g. positive emotion, high attention and positive trend of game performances). In that case, the robot can react appropriately and enthusiastically to maintain the user's engagement. Alternatively, if the user appears in a negative state, the robot should use inciting behaviour to encourage the user to remain attentive and attempt to re-engage with the robot. The

adaptation policy will dictate the appropriate course of action for the robot based on the user's current state during a serious game.

2.1.1. Cooking Serious Game Scenario.

The cooking game involves eight questions that challenge users to identify the correct sequence and weight of the ingredients. The serious game is divided into five stages: introduction, recipe instruction, question state, answer state and ending feedback. At the start of the application, the robot greets the user and asks if they are prepared to play. When the cooking game starts, the robot shows and vocally synthesises the ingredients for the selected recipe. The robot emphasises the sequential ingredients' order and weight during the recipe instruction. The quizzes follow, in which the user must use visual attention and working memory to identify the correct ingredients and choose them from the available options. The user interacts with the game using voice modality. The cooking game topic was chosen considering the previous experiences of the psychologists and psychotherapist experts in cognitive training with MCI users [14, 22] and a literature analysis [23, 24]. After a semi-structured interview with three psychologists and one neuroscience researcher of CNR Pisa, it was decided to design a cooking application connected to the user's daily activity, requiring users to recognise the ingredients' chronological sequence, weight, and typology of ingredients.

2.1.2. Robot Personality.

The robot in our scenario is a Pepper robot and exploits two personalities: extraverted and introverted one [25]. Typically, extroverts tend to speak in a louder, faster, and higher-pitched manner. They are also more inclined to initiate conversations and talk more about themselves than others. Regarding body language, their gestures and movements are generally more expansive and faster and occur more frequently than introverted ones. Conversely, for the introverted condition, the robot's gestures tend to be more limited, contained, and slower in such a way to appear reserved toward the user.

2.1.3. Q-learning.

In the Q-learning algorithm, interactions consist of a sequence of user states (S), robot actions (A), and rewards (R) (see Figure 1). The RL agent observes the state S_t from the environment and chooses an action A_t to perform based on this observation. The chosen action is then executed, and the environment provides a new state, S_{t+1} , and a reward, R_{t+1} , to evaluate the transition. Given the current state, the agent's policy (Q -table) maps states to actions and determines which action to select.

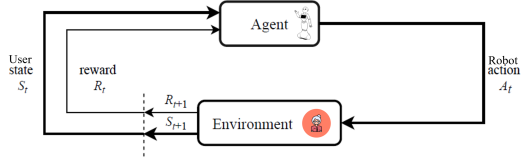


Figure 1: Reinforcement learning standard framework applied to HRI scenario.

2.1.4. Fuzzy Logic System.

Fuzzy logic is a mathematical approach to emulate the human way of thinking and learning [26]. A Fuzzy Logic systems depend on prior expert knowledge to specify the fuzzy sets and fuzzy rule bases. A fuzzy logic system consists of three components: fuzzification, fuzzy logic controller, and defuzzification [20]. The first component converts the crisp values (i.e. not fuzzy values) of the input variable that defined the user state (user’s emotion, attention, and game performance) into their fuzzy form using some membership functions. A membership function is a curve that defines how each point in the input space is mapped to a membership value (or degree of membership) between 0 and 1. The input variables are assigned to the linguistic variable (i.e. variables with a value of linguistic concepts rather than numbers, e.g. High attention, Low attention, Intermediate attention). The fuzzy output of each input variable can take one candidate value from a set of defined values [27]. In our simulations, three user models were created (one for healthy users, one for MCI, and one for users with dementia) to tailor the adaptation. The user model varies according to the cognitive level of the user, the user’s performance and the user’s attention.

The second one simulates human reasoning by making fuzzy inferences based on inputs and a pre-defined set of IF-THEN rules. The rules allow us to define the relationship between the input (state) and the output (action). We define the rules using our prior knowledge based on our experience with MCI and an experiment with robot adaptation behaviour using Q-learning. The fuzzy logic component’s output is an algebraic product of the degree of truth of each fuzzy input defined in the fuzzification phase. The last component helps to convert the fuzzy output set from the linguistic variables into a crisp value. The fuzzy output of the system can take one value from a set of 7 constant values from extremely low to extremely high in a range between i.e. [-5,+5].

In summary, Fuzzy Logic resembles the human decision-making methodology and deals with vague and imprecise information. It is an approach to computing based on "degrees of truth" rather than the usual "true or false" (1 or 0) Boolean logic. According to our goal

for an adaptive robot behaviour generation that maintains a positive user state through interaction with the robot, we define the key elements of the Fuzzy Q-learning algorithm, described in the following section.

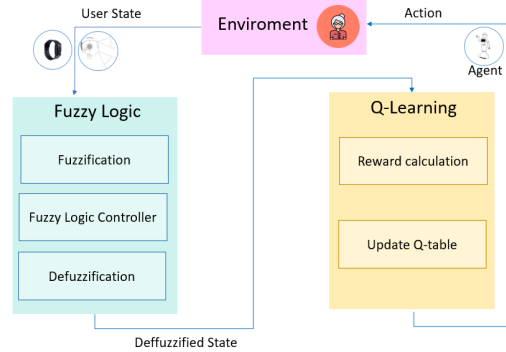


Figure 2: Fuzzy Reinforcement learning model proposed.

2.2. Modelling an adaptive Fuzzy Q-learning application

In our Fuzzy Q-learning-based model proposal, the user state is considered the environment, while the robot is the agent (see Figure 2).

State. Three variables describe the user’s state: the user’s emotion, the user’s attention, and the user’s game performance. The user emotions considered are seven basic emotions: happiness, neutral, sad, surprise, fear, disgust, and anger, identified using physiological parameters collected by the Empatica E4 wristband [28]. The user’s attention is extrapolated, collecting the user’s gaze direction using the QisdK library of the Pepper robot and the user’s performance considering the trend of correct and wrong answers during the session with the robot [29]. All these environment information that identified the user state are fuzzified. The triangular and trapezoidal membership functions are used to fuzzify the parameters and identify the linguistic variables: high, low, and intermediate user states that vary from extremely low to extremely high user-level states.

Action. The actions identified correspond to the behavioural responses exhibited by the robot, encompassing verbal feedback, vocal characteristics, animations, and motor movements tailored to the supported robot personalities. The robot can execute three distinct actions within the serious gaming context. Firstly, (a0) entails generating an enthusiastic, more spirited behaviour. For instance, the robot manifests enthusiastic feedback within the extravert personality paradigm, exemplified by phrases such as "My gosh! That is the correct one!"

You are trying hard!" This is accompanied by a slight augmentation in speech rate, volume, and pitch, alongside dynamic and expansive animations featuring pronounced motor movements. Secondly, (a1) pertains to generating a more inciting behaviour, typified by phrases like "right answer! Let us continue with this focus!" and animations characterized by closer proximity. Finally, (a2) involves the generation of a more neutral robotic comportment. An instance is provided by the utterance "Good! That is the right answer!" delivered with vocal parameters reflecting neutrality as per the robot's personality configuration, alongside basic animations consonant with its inherent persona.

Reward The robot must avoid the user's descent into negative states (e.g. negative emotion, negative performance and low attention). The selection of the reward weights aims to optimise the robot behaviour policy, thereby maintaining the user in a positive state while facilitating a positive trajectory in performance. This becomes particularly salient in repetitive cognitive training, where task monotony prevails. Emphasizing a more authentic and captivating robot behaviour centred around the user's state and performance is important in sustaining user engagement and active participation within this context. Tailoring more adapting robot behaviours for this demographic assumes significance, as it enhances user acceptability. Consequently, users are afforded a heightened sense of comfort and focus during training sessions, fostering a more enjoyable and effective interaction facilitated by incorporating human-like behaviour or actions within the robot's repertoire.

The system illustrated in Figure 2 will be tested in a simulated environment using Python with the three-user model defined and then implemented in the Pepper robot for a real-user test.

3. Robot Adaptive Application

The humanoid robot that will be employed for the real user test is the Pepper model, developed by Softbank's Robotics. Pepper is 1.2 meters tall and has 17 joints to facilitate expressive body language, alongside three omnidirectional wheels to facilitate mobility. With a suite of multimodal interfaces, including touchscreen, speech, tactile head, hands, bumper, LEDs, and 20 degrees of freedom for whole-body motion, Pepper offers a versatile platform for interaction. In this proposal, the camera sensors are leveraged to detect various user attention states, in particular, by categorizing the gaze direction. Gaze direction is categorized into five values: (1) direct eye contact with the robot, (2) looking upward, (3) focusing on the tablet, (4) looking left, and (5) looking right. The user is deemed to be in an "attention state" when directing their gaze toward the robot or

the tablet, while other directions denote a "distracted state". These defined parameters are captured during interactions with the robot using modules provided by the QiSDK robot framework. Data collection intervals are set at every one second to ensure an adequate dataset for assessing attentive states[29]. Regarding emotion recognition, the Empatica E4 wristband will be used. The emphatic E4 has four sensors, including an Electrodermal activity sensor (μS), a heart rate sensor (beats per minute), skin temperature (C), a 3-axis accelerometer on three orthogonal axes (X, Y, Z), and an optical thermometer[28]. The E4 wristband collected the heart rate at a frequency of 1 Hz, the skin temperature and EDA at 4 Hz, and the acceleration at 32 Hz[28]. Regarding emotion classification, various machine learning algorithms were used to recognize emotion using physiological data. A possible suitable algorithm for emotion recognition identified by literature could be the Vector Machine (SVM), supervised learning algorithms that offer promising results in high emotion recognition accuracy [30, 31]. Regarding the trend of user performance during the gameplay, the serious game collects different values such as the number of correct answers, the number of tentative answers right, reaction time, session time, and other parameters related to the game. Following the European regulation about privacy and data, these are saved into a CNR database. The robot and wristband will determine the user's current state, combining these values with the game data collected by the serious game implemented in Android. In parallel, the Android application manages the robot's data sensing and the wristband data collection. In contrast, the Fuzzy logic module and the emotion recognition algorithm will be implemented in Python.

After the training phase, which simulates the fuzzy logic system with the three user profiles defined, a real user test will be held in a laboratory setting at the CNR of Pisa. The users will test a robot in a within-subject study design in a random condition with an adaptive condition. The goal of the test will be to identify if the user can perceive an adaptation of the robot's behaviour concerning the random condition and if the User Engagement in the adaptation condition is higher with respect to the random one. To evaluate these hypotheses, the user will compile the User Engagement Scale [32], and Godspeed questionnaires [33] at the end of each robot condition interaction.

4. Conclusions

This paper describes a proposal of a Fuzzy RL algorithm to learn the best policy strategy in a SAR running two personalities that help to cope in the context of cognitive

training individuals with MCI, healthy and with dementia. The proposed model allows the robot's behavioural system to select an appropriate action based on the state of the fuzzified user. Fuzzy logic enables the circumvention of the constraints inherent in Q-learning by facilitating the selection of a set of overarching and generic states. In the proposed model, the possible actions that the robot performs to stimulate have been identified to increase user engagement during a serious gaming scenario. The designated actions follow the parameters defined for the extroverted and introverted personalities introduced in section 2.1.2. In future work, we want to implement the proposed Fuzzy Q-learning model and validate it by evaluating the metrics, such as average reward, average step to complete the task, and average computation time in the simulation phase[18] and then test the system with real users. Another step is identifying the best machine-learning algorithm for classifying user emotions collected by the Empatica wristband. In conclusion, this article describes a proposal to provide an adaptation policy for robot actions that adapt to fuzzy user states.

References

- [1] F. Carros, J. Meurer, D. Löffler, D. Unbehaun, S. Matthies, I. Koch, R. Wieching, D. Randall, M. Hassenzahl, V. Wulf, Exploring human-robot interaction with the elderly: results from a ten-week case study in a care home, in: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 2020, pp. 1–12.
- [2] N. Akalin, A. Loutfi, Reinforcement learning approaches in social robotics, *Sensors* 21 (2021) 1292.
- [3] S. Keizer, M. Ellen Foster, Z. Wang, O. Lemon, Machine learning for social multiparty human-robot interaction, *ACM transactions on interactive intelligent systems (TIIS)* 4 (2014) 1–32.
- [4] J. De Greeff, T. Belpaeme, Why robots should be social: Enhancing machine learning through social human-robot interaction, *PLoS one* 10 (2015) e0138061.
- [5] J. Hemminghaus, S. Kopp, Towards adaptive social behavior generation for assistive robots using reinforcement learning, in: Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, 2017, pp. 332–340.
- [6] K. Tsiakas, M. Dagioglou, V. Karkaletsis, F. Makeidon, Adaptive robot assisted therapy using interactive reinforcement learning, in: International Conference on Social Robotics, Springer, 2016, pp. 11–21.
- [7] H. Ritschel, A. Seiderer, K. Janowski, S. Wagner, E. André, Adaptive linguistic style for an assistive robotic health companion based on explicit human feedback, in: Proceedings of the 12th ACM international conference on Pervasive technologies related to assistive environments, 2019, pp. 247–255.
- [8] C. Pou-Prom, S. Raimondo, F. Rudzicz, A conversational robot for older adults with alzheimer's disease, *ACM Transactions on Human-Robot Interaction (THRI)* 9 (2020) 1–25.
- [9] L.-F. Chen, Z.-T. Liu, M. Wu, F.-Y. Dong, Y. Yamazaki, K. Hirota, Multi-robot behavior adaptation to local and global communication atmosphere in humans-robots interaction, *Journal on Multimodal User Interfaces* 8 (2014) 289–303.
- [10] A. Tapus, Improving the quality of life of people with dementia through the use of socially assistive robots, in: 2009 Advanced Technologies for Enhanced Quality of Life, IEEE, 2009, pp. 81–86.
- [11] J. Magyar, M. Kobayashi, S. Nishio, P. Sinčák, H. Ishiguro, Autonomous robotic dialogue system with reinforcement learning for elderlies with dementia, in: 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), IEEE, 2019, pp. 3416–3421.
- [12] H. Modares, I. Ranatunga, F. L. Lewis, D. O. Popa, Optimized assistive human-robot interaction using reinforcement learning, *IEEE transactions on cybernetics* 46 (2015) 655–667.
- [13] E. Zedda, M. Manca, F. Paterno, C. Santoro, Mci older adults' user experience with introverted and extraverted humanoid robot personalities, in: Proceedings of the 15th Biannual Conference of the Italian SIGCHI Chapter, 2023, pp. 1–12.
- [14] M. Manca, F. Paternò, C. Santoro, E. Zedda, C. Braschi, R. Franco, A. Sale, The impact of serious games with humanoid robots on mild cognitive impairment older adults, *International Journal of Human-Computer Studies* 145 (2021) 102509.
- [15] R. S. Sutton, Learning to predict by the methods of temporal differences, *Machine learning* 3 (1988) 9–44.
- [16] R. S. Sutton, A. G. Barto, Reinforcement learning: An introduction, MIT press, 2018.
- [17] C. J. Watkins, P. Dayan, Q-learning, *Machine learning* 8 (1992) 279–292.
- [18] A. Bonarini, Evolutionary learning, reinforcement learning, and fuzzy rules for knowledge acquisition in agent-based systems, *Proceedings of the IEEE* 89 (2001) 1334–1346.
- [19] R. E. Bellman, L. A. Zadeh, Decision-making in a fuzzy environment, *Management science* 17 (1970) B–141.
- [20] A. Bonarini, A. Lazaric, F. Montrone, M. Restelli, Reinforcement distribution in fuzzy q-learning, *Fuzzy sets and systems* 160 (2009) 1420–1443.
- [21] L. Zadeh, Fuzzy sets, *Inform Control* 8 (1965) 338–353.

- [22] E. Zedda, M. Manca, F. Paternò, A cooking game for cognitive training of older adults interacting with a humanoid robot., in: CHIRA, 2021, pp. 271–282.
- [23] V. Manera, P.-D. Petit, A. Derreumaux, I. Orvieto, M. Romagnoli, G. Lyttle, R. David, P. H. Robert, ‘kitchen and cooking,’ a serious game for mild cognitive impairment and alzheimer’s disease: a pilot study, *Frontiers in aging neuroscience* 7 (2015) 24.
- [24] S. McCallum, C. Boletsis, Dementia games: A literature review of dementia-related serious games, in: *Serious Games Development and Applications: 4th International Conference, SGDA 2013, Trondheim, Norway, September 25-27, 2013. Proceedings* 4, Springer, 2013, pp. 15–27.
- [25] E. Zedda, M. Manca, F. Paternò, C. Santoro, Older adults’ user experience with introvert and extravert humanoid robot personalities, *Universal Access in the Information Society* (2023) 1–17.
- [26] R. Sharma, M. Gopal, A markov game-adaptive fuzzy controller for robot manipulators, *IEEE Transactions on Fuzzy Systems* 16 (2008) 171–186.
- [27] R. D. Hegazy, O. A. Nasr, A user behavior based handover optimization algorithm for lte networks, in: *2015 IEEE Wireless Communications and Networking Conference (WCNC)*, IEEE, 2015, pp. 1255–1260.
- [28] E. Inc, E4wristband: Real-time physiological signals: Wearable ppg, eda, temperature, motion sensors., *Empatica.com* (last visit: 12/14/2024). URL: <https://www.empatica.com/en-eu/research/e4/>.
- [29] Softbank Robotics, Human, QiSDK (2019). URL: https://qisdk.softbankrobotics.com/sdk/doc/pepper-sdk/ch4_api/perception/reference/human.html#retrieving-characteristics.
- [30] M. Ragot, N. Martin, S. Em, N. Pallamin, J.-M. Di-verrez, Emotion recognition using physiological signals: laboratory vs. wearable sensors, in: *Advances in Human Factors in Wearable Technologies and Game Design: Proceedings of the AHFE 2017 International Conference on Advances in Human Factors and Wearable Technologies*, July 17-21, 2017, The Westin Bonaventure Hotel, Los Angeles, California, USA 8, Springer, 2018, pp. 15–22.
- [31] L. Shu, J. Xie, M. Yang, Z. Li, Z. Li, D. Liao, X. Xu, X. Yang, A review of emotion recognition using physiological signals, *Sensors* 18 (2018) 2074.
- [32] H. L. O’Brien, P. Cairns, M. Hall, A practical approach to measuring user engagement with the refined user engagement scale (ues) and new ues short form, *International Journal of Human-Computer Studies* 112 (2018) 28–39.
- [33] C. Bartneck, D. Kulić, E. Croft, S. Zoghbi, Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots, *International journal of social robotics* 1 (2009) 71–81.