Building a Better Oracle: Using Personas to Create More Human-Like Oracles

Jindan Huang
Tufts University
jindan.huang@tufts.edu

Elaine Schaertl Short
Tufts University
elaine.short@tufts.edu

KEYWORDS

Human-Robot Interaction; Interactive Reinforcement Learning; Participatory Design; Persona

1 INTRODUCTION

With tremendous potential to improve quality of life, collaborative and assistive robots are increasingly used in various scenarios, such as education [1] and healthcare [2]. In order to achieve effective and efficient human-robot collaboration, these robots must be able to adapt to dynamic environments and changing demands from different users.

One technique that robotic researchers commonly use is *Interactive Reinforcement Learning*, which allows agents to use human feedback in order to learn faster. However, most prior work is tested with perfect oracles that always knows the optimal policy for the current state and immediately provides feedback to the learning agent. However, in real human-robot interaction scenes, human teachers could give feedback in a delayed, stochastic and unsustainable ways [3]. Also, humans' feedback on the same thing may vary because of their different personalities, preferences and experience. Because of this, a perfect oracle is not an accurate reflection of human teaching, and developing algorithms with such oracles is not likely to result in approaches that work with real users in real-world robotics problems.

Regarding the aforementioned issues, modifying an unrealistic perfect oracle to a more human-like one is a helpful remedy. In this paper, we propose a Research through Design (RtD) workflow by introducing the concept of Participatory Design and Persona into interactive reinforcement learning algorithm design. Together with participants, we will create a set of representative persona models for different types of users. Then, exposed to a simulated robot environment with a persona-based oracle, participants can think aloud and adjust the oracle's feedback behaviors through a UI control interface. Later, by analyzing user data from the participatory design sessions, we can generate a set of modified oracles with respect to user types.

Instead of directly transforming persona features into task-specific rewards, our proposed method builds connections between each persona and its corresponding feedback behaviors. The modified oracles we generate can be used for other researchers to evaluate how the performance of their algorithms varies with different user profiles and to gain inspiration for improving the algorithms.

2 BACKGROUND

Interactive reinforcement learning, as a branch of reinforcement learning (RL), involves a human-in-the-loop that tailors specific elements of the underlying RL algorithm to improve its performance or produce an appropriate policy for a particular task. Compared to traditional RL algorithms, interactive RL better integrates human feedback and has proven to be more effective on agent learning [4–6]. Furthermore, learning rates can be significantly improved and the number of required trials can be reduced, if RL algorithms are trained with both prior knowledge and human feedback [7].

The societal nature of Human-Robot Interaction (HRI) determines the necessity of thinking from all stakeholders' perspectives when applying interactive RL to HRI research. However, one gap is that researchers have yet to propose a clear solution to design efficient interactive RL methods that can be adapted to feedback from various user types, given that it is rarely feasible to do user testing at every iteration of algorithm development [8]. Prior work has found that variations in human feedback could occur in different aspects such as frequency [9] and strategy [10]. Therefore, to fill the gap mentioned above, in-depth research of human feedback and the reasons resulting in its change is a must.

One of the most common methodologies to better understand user behaviors is Participatory Design [11]. Rather than prescribing solutions to people based on designer's assumptions, participatory design emphasizes on user-designer collaboration and aims to integrate professional viewpoints with existing user knowledge, which would reduce the risk of design failure and foster the exploration of new solutions.

With extensive quantitative and qualitative user data from participatory design, the technique of Persona [12] is needed to synthesize all results by representing a group of users with similar personalities and preferences using a vivid, fictional character. Though few RL research studies have incorporated this technique, it has been proved to be effective for maintaining personality consistency [13], prompting more engaged conversation [14, 15] and generating personalized robot behaviors [16].

3 METHODOLOGY

3.1 Persona-Feedback Model

The Persona-Feedback model focuses on integrating the persona methodology to modify a perfect oracle's feedback(see Figure 1). It aims to build a relationship mapping between personas and feedback behaviors. In order to construct the conceptual structure of this model, the following questions need to be answered:

- Which are the key persona variables we should consider?
- Why can these variables cause differences in feedback?
- What aspects of feedback behaviors will be changed?

3.1.1 Persona Variables. The persona definition must at least include three variables: demographic information(age, gender, education, etc.), personality and technical expertise. These variables are found to be influential in interacting with robots [17] and associated with evaluations of the agent's behaviors [18]. Depending on the research setting, other optional variables, such as mental state and physical reactivity, could be included as well.

3.1.2 Intermediate Reasoning. As persona variables could directly or indirectly affect feedback behaviors, we introduce intermediate reasoning into our model to better understand the influence process. The following list contains six latent variables that should be considered, although more may exist, especially in specialized application domains.

- `Attention`: in relation to the delay when giving feedback
- `Reaction Ability`: in relation to find the user's feedback frequency as well as feedback delay
- `Errors`: whether the feedback reflects the performance of the agent in a proper way, in relation to the accuracy of the user's feedback relative to the underlying reward function
- `Expectation`: the user's expectations for the agent's behavior, in relation to secondary goals that the agent should achieve, such as avoiding mistakes, exploiting known solutions more, etc.
- `Affection`: how much the user likes the agent in relation to how much positive feedback the user will give
- `Confidence`: the user's confidence level when giving feedback, in relation to how much the agent should weight that feedback

3.1.3 Variations in feedback. Unlike a perfect oracle, human teachers are not able to give unchanging and precise numerical rewards without delay or at every time step [3]. Besides, feedback strategies could vary between individuals [10]. Thus, in Table 1, we present attributes which may cause variations in feedback.

3.2 Procedure

Building on this framework for the Persona-Feedback model, we use participatory design methodology to refine its details.

3.2.1 Preliminary Preparation. Upon consenting to participate in this study, participants will complete a brief background survey covering their demographic information(age, gender, area of study, occupation) and previous technical experience(level of programming, familiarity with robots). The Big Five Personality Test could be added for specific research purposes. Then, we will send each participant a private Zoom link for a one-on-one workshop. We will record the whole meeting to collect the participant's facial expressions, gestures and verbal reactions.

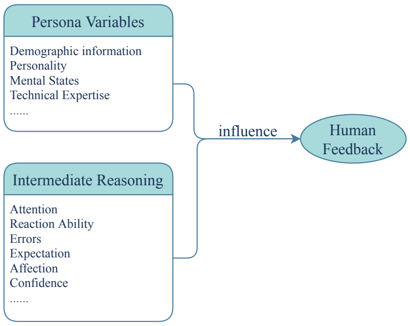
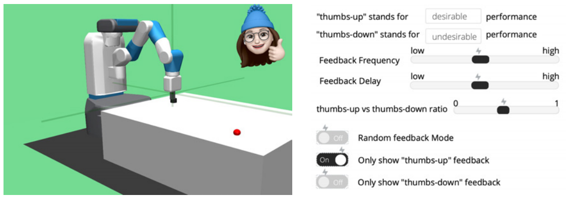


Figure 1: The Principle of Persona-Feedback Model



"thumbs-up" stands for desirable performance
"thumbs-down" stands for undesirable performance
Feedback Frequency low high
Feedback Delay low high
thumbs-up vs thumbs-down ratio 0 1
○ Off Random feedback Mode
● On Only show "thumbs-up" feedback
○ Off Only show "thumbs-down" feedback

Figure 2: (left) An example of simulated robot environment with a persona-based oracle giving feedback; (right) a mock UI control interface

3.2.2 Persona Development Session. The first step will be a persona development session, where a participant works with a researcher to create two personas: one representing themselves and another for someone they know well. Without explicitly presenting a persona example, the researcher will give an initial guideline as described above. Depending on the participant, other characteristics including mood, profession, dis/ability status, or other identities could be added on a more ad-hoc basis. Participants will also have to choose a portrait from a free stock photo library, serving as the avatar of each persona. In addition, we always keep the age of the persona ambiguous, for example, referring to it as "someone in his 30s". This was done as to keep the persona fluid between participants and also let the participant easily come up with their own perception.

To gain intermediate reasoning for further feedback behavior design and avoid any stereotype affecting persona development, we will prompt participants to share their personal stories and answer some situational questions. Below are some sample questions.

- Have you ever been a tutor/teaching assistant and can you share your experience?
- Imagine if you are teaching a robot, what kind of teaching style will you use?

3.2.3 Co-design Session. The second step will be a co-design session. The participants will be asked to observe a simulated environment(see Figure 2 left), where the persona avatars designed in the previous session will be shown evaluating the performance of a robot agent. The agent will be completing a task such as reaching a point or maintaining the balance, and initially the avatar gives feedback as a perfect oracle. Since requesting people give numerical scores is known to be difficult [19], feedback from the avatar will be provided as a binary form of either "thumbs-up" or "thumbs-down". Then, each participant is allowed to change the feedback pattern though a UI control interface(see Figure 2 right) in order to fit the avatar's personality. To maintain consistency, we will review persona characteristics of the avatar together with participants and make sure they empathize with the persona's perspective.

On the UI control interface, participants can define the meanings of feedback indicators, as well as choose to display only "thumbs-up" or "thumbs-down" by switching corresponding toggles. The value of feedback frequency, delay time, and the ratio of two different kinds of feedback can also be adjusted through a draggable bar. When participants are satisfied with his modification, we will record the parameter value of each feedback attribute. After this session, we should get two sets of feedback behavior parameters from each participant, since every individual designs two personas.

Table 1: Changeable Feedback Attributes

Attribute	Definition	Categorical/Numerical Description
`Delay`	the time that the teacher takes to give feedback	response time
`Frequency`	how often the teacher gives feedback	the number of time steps covered in one feedback
`Correlation`	the relationship between feedback and long-term reward	*positively-correlated*, *anti-correlated*, or *uncorrelated*
`Motivation`	whether feedback focuses on positive reinforcement or punishment	*reward-focused*, *punishment-focused*, or *balanced*
`Bias`	whether feedback is overall more negative or positive	a ratio between 0 and 1
`Importance`	how much the feedback should affect the agent	a weighing factor between 0 and 1

3.2.4 Evaluation Session. Finally, we will recruit new participants for an evaluation and think aloud session. Participants will be provided with any two of persona models designed in the previous session. For each persona, the participants will watch two short videos about the same simulated scenarios used in the previous co-design session, however, using two different feedback patterns: 1) a perfect oracle style and 2) a human-preferred style. The human-preferred style comes from the feedback attribute data we collected in the previous session. In each video, a persona is present to give feedback for performance of the learning agent with one feedback pattern. After each video, participants have to answer the following open-ended and rating scale questions, where 1 is the least and 5 is the most:

- How realistic is this persona on a scale from 1 to 5?
- To what degree do you think the feedback behaviors in this video match the persona on a scale from 1 to 5?
- Where could be changed to better match this persona?

4 DISCUSSION

Contributions and Implications for HRI community. In this paper, we present a participatory design workflow that generates human-like feedback behaviors for reinforcement learning oracles. The main contribution is that we expand the practice of RtD approach to HRI communities and enable the development of more robust interactive RL algorithms. Our work solves possible problems caused by the flawed assumption of a perfect oracle, integrating the characteristics of real human feedback behaviors into the robot learning process. By introducing the concept of persona, we also eliminate the requirement of endless user involvement. For other HRI researchers, our idea brings new opportunities to evaluate the performance of their RL algorithms within different user groups, and is beneficial to gain inspirations for further optimization.

Reflection and Future Work. The limitation of our work could exist in the persona development process. We should avoid cultural stereotypes interfering the quality of our persona models. Also, we realize that it is important to design an accessible RtD procedure for participants, since many of them could be non-experts in robotics or computer science. We should always make sure participants can easily understand and communicate with researchers without having prior professional knowledge. The next step will be to evaluate and validate our methodology by conducting several case studies within different focus groups. Finally, we will release our personas and their corresponding parameters of feedback attributes for other HRI researchers to use in their own work.

REFERENCES

[1] Elaine Short, Katelyn Swift-Spong, Jillian Greczek, Aditi Ramachandran, Alexandru Litoiu, Elena Corina Grigore, David Feil-Seifer, Samuel Shuster, Jin Joo Lee, Shaobo Huang, et al. How to train your dragonbot: Socially assistive robots for teaching children about nutrition through play. In *The 23rd IEEE international symposium on robot and human interactive communication*, pages 924–929. IEEE, 2014.

[2] Reza Kachouie, Sima Sedighadeli, Rajiv Khosla, and Mei-Tai Chu. Socially assistive robots in elderly care: a mixed-method systematic literature review. *International Journal of Human-Computer Interaction*, 30(5):369–393, 2014.

[3] Riku Arakawa, Sosuke Kobayashi, Yuya Unno, Yuta Tsuboi, and Shin-ichi Maeda. Dqn-tamer: Human-in-the-loop reinforcement learning with intractable feedback. *arXiv preprint arXiv:1810.11748*, 2018.

[4] Jerry Alan Fails and Dan R Olsen Jr. Interactive machine learning. In *Proceedings of the 8th international conference on Intelligent user interfaces*, pages 39–45, 2003.

[5] Shane Griffith, Kaushik Subramanian, Jonathan Scholz, Charles L Isbell, and Andrea L Thomaz. Policy shaping: Integrating human feedback with reinforcement learning. Georgia Institute of Technology, 2013.

[6] Garrett Warnell, Nicholas Waytowich, Vernon Lawhern, and Peter Stone. Deep tamer: Interactive agent shaping in high-dimensional state spaces. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

[7] Rachit Dubey, Pulkit Agrawal, Deepak Pathak, Thomas L Griffiths, and Alexei A Efros. Investigating human priors for playing video games. *arXiv preprint arXiv:1802.10217*, 2018.

[8] Christian Arzate Cruz and Takeo Igarashi. A survey on interactive reinforcement learning: Design principles and open challenges. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, pages 1195–1209, 2020.

[9] Charles Isbell, Christian R Shelton, Michael Kearns, Satinder Singh, and Peter Stone. A social reinforcement learning agent. In *Proceedings of the fifth international conference on Autonomous agents*, pages 377–384, 2001.

[10] Matthew E Taylor and AI Borealis. Improving reinforcement learning with human input. In *IJCAI*, pages 5724–5728, 2018.

[11] Michael J Muller and Sarah Kuhn. Participatory design. *Communications of the ACM*, 36(6):24–28, 1993.

[12] Alan Cooper, Robert Reimann, et al. *About face 2.0: The essentials of interaction design*, volume 17. Wiley Indianapolis, 2003.

[13] Jiwei Li, Michel Galley, Chris Brockett, Georgios P Spithourakis, Jianfeng Gao, and Bill Dolan. A persona-based neural conversation model. *arXiv preprint arXiv:1603.06155*, 2016.

[14] Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, and Jason Weston. Personalizing dialogue agents: I have a dog, do you have pets too? *arXiv preprint arXiv:1801.07243*, 2018.

[15] Mohsen Mesgar, Edwin Simpson, Yue Wang, and Iryna Gurevych. Generating persona-consistent dialogue responses using deep reinforcement learning. *arXiv preprint arXiv:2005.00036*, 2020.

[16] Antonio Andriella, Carme Torras, and Guillem Alenyà. Learning robot policies using a high-level abstraction persona-behaviour simulator. In *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1–8. IEEE, 2019.

[17] Ismael Duque, Kerstin Dautenhahn, Kheng Lee Koay, Bruce Christianson, et al. A different approach of using personas in human-robot interaction: Integrating personas as computational models to modify robot companions' behaviour. In *2013 IEEE RO-MAN*, pages 424–429. IEEE, 2013.

[18] Dag Sverre Syrdal, Kerstin Dautenhahn, Kheng Lee Koay, and Michael L Walters. The negative attitudes towards robots scale and reactions to robot behaviour in a live human-robot interaction study. *Adaptive and emergent behaviour and complex systems*, 2009.

[19] Carolyn C Preston and Andrew M Colman. Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta psychologica*, 104(1):1–15, 2000.