Facial Feedback for Reinforcement Learning: A Case Study and Offline Analysis Using the TAMER Framework

JAAMAS Track

Guangliang Li
Ocean University of China
Qingdao, China
guangliangli@ouc.edu.cn

Hamdi Dibeklioğlu
Bilkent University
Ankara, Turkey
dibeklioglu@cs.bilkent.edu.tr

Shimon Whiteson
University of Oxford
Oxford, UK
shimon.whiteson@cs.ox.ac.uk

Hayley Hung
Delft University of Technology
Delft, The Netherlands
h.hung@tudelft.nl

ABSTRACT

Interactive reinforcement learning provides a way for agents to learn to solve tasks from evaluative feedback provided by a human user. Previous research showed that humans give copious feedback early in training but very sparsely thereafter. In this paper, we investigate the potential of agent learning from trainers' facial expressions via interpreting them as evaluative feedback. To do so, we implemented TAMER which is a popular interactive reinforcement learning method in a reinforcement-learning benchmark problem — Infinite Mario, and conducted the first large-scale study of TAMER involving 561 participants. With designed CNN-RNN model, our analysis shows that telling trainers to use facial expressions and competition can improve the accuracies for estimating positive and negative feedback using facial expressions. In addition, our results with a simulation experiment show that learning solely from predicted feedback based on facial expressions is possible and using strong/effective prediction models or a regression method, facial responses would significantly improve the performance of agents. Furthermore, our experiment supports previous studies demonstrating the importance of bi-directional feedback and competitive elements in the training interface.

KEYWORDS

Facial Feedback, Interactive Reinforcement Learning, Implicit Feedback

ACM Reference Format:
Guangliang Li, Hamdi Dibeklioğlu, Shimon Whiteson, and Hayley Hung. 2021. Facial Feedback for Reinforcement Learning: A Case Study and Offline Analysis Using the TAMER Framework: JAAMAS Track. In *Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), Online, May 3–7, 2021*, IFAAMAS, 3 pages.

1 INTRODUCTION

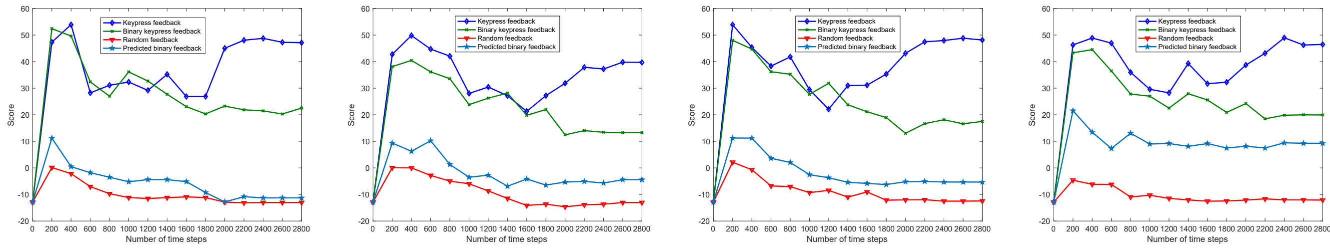
With the recent advances in AI, socially intelligent autonomous agents are becoming our high-tech companions in the family. The ability of these intelligent agents to efficiently learn from non-technical users to perform a task in a natural way will be key to their success. Therefore, it is critical to develop methods that facilitate the interaction between these non-technical users and agents, through which they can transfer task knowledge effectively to such agents.

Interactive reinforcement learning has proven to be a powerful technique for facilitating the teaching of artificial agents by their human users [1, 2, 4, 5, 7, 9–11, 13]. In interactive reinforcement learning, an agent learns from human reward, i.e., evaluations of the quality of the agent's behavior provided by a human user, in a reinforcement learning framework. Nonetheless, agent learning from human reward is limited by the quality of the interaction between the human trainer and agent. Several TAMER studies — a popular interactive reinforcement learning method for enabling autonomous agents to learn from human reward [4], have shown that humans give copious feedback early in training but very sparsely thereafter [3, 6, 8]. As facial expressions have been often used by humans to consciously or subconsciously encourage or discourage specific behaviors they want to teach [12], we investigate the potential of using facial expressions as reward signals in our study.

To examine this potential, we conducted the first large-scale study of TAMER by implementing it in the Infinite Mario domain. Our study, involving 561 participants, at the NEMO science museum in Amsterdam using museum visitors (aged 6 to 72). We recorded the facial expressions of all trainers during training and, in some conditions, told participants that their facial expressions would be used as encouraging explicit feedback, e.g., happy and sad expressions would map to positive and negative reward respectively, in addition to keypresses, to train the agent.

The experimental results show that telling trainers to use facial expressions makes them inclined to exaggerate their expressions, resulting in higher accuracies for estimating their corresponding positive and negative feedback keypresses using facial expressions. Moreover, competition can also elevate facial expressiveness and further increase the predicted accuracy. Furthermore, with designed CNN-RNN model, our results in a simulation experiment show that it is possible for an agent to learn solely from predicted evaluative feedback based on facial expressions. To our knowledge, it is the first time facial expressions have been shown to work in TAMER,

Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), U. Endriss, A. Nowé, F. Dignum, A. Lomuscio (eds.), May 3–7, 2021, Online.
© 2021 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.



(a) Control condition (b) FE condition (c) Competitive condition (d) Competitive FE condition

Figure 1: Offline performance of agent learning from predicted binary feedback with facial expressions, compared to learning from keypress feedback, binary keypress feedback and random feedback for all four conditions.

opening the door to a much greater potential for learning from human reward in more natural, personalized and possibly more long term learning scenarios.

Table 1: Accuracy of classifying positive and negative feedback using facial responses.

	Condition	Positive	Negative	Total
Proposed Method	Control	0.62	0.69	0.66
	Facial Expression	0.65	0.73	0.70
	Competitive	0.75	0.70	0.73
	Competitive Facial Expression	0.79	0.75	0.78
Random Baseline	Control	0.50	0.50	0.50
	Facial Expression	0.42	0.58	0.51
	Competitive	0.52	0.48	0.50
	Competitive Facial Expression	0.58	0.42	0.51

2 EXPERIMENTAL CONDITIONS AND SETUP

We investigate how ‘facial expression’ and ‘competition’ affect the agent’s learning performance and trainer’s facial expressiveness in four experimental conditions: the *control condition*—without ‘competition’ or ‘facial expression’, the *facial expression condition*—without ‘competition’ but with ‘facial expression’, the *competitive condition*—with ‘competition’ but without ‘facial expression’, and the *competitive facial expression condition*—with both. We hypothesize that ‘competition’ will result in better performing agents, and ‘facial expression’ will result in worse agent performance.

Our experiment is a between-subjects study with 561 participants from more than 27 countries and randomly distributed into our four experimental conditions. Of them, 221 were female and 340 were male respectively, aged from 6 to 72. After pruning the data, 498 participants remained: 109 participants in the control condition; 100 in the facial expression condition; 135 in the competitive condition; and 154 in the competitive facial expression condition.

3 EXPERIMENTAL RESULTS AND ANALYSIS

We analyzed the predicted accuracy of implicit feedback based on the recorded facial expression data and tested the learning performance from implicit facial feedback.

3.1 Classification of Positive and Negative Feedback with Facial Expressions

We designed and trained a CNN-RNN model with recorded data to predict feedback with facial expressions. A random baseline is also reported for comparison, as shown in Table 1. Class labels for random baseline are assigned by drawing a random class label according to the ratio of positive and negative class labels from the training set. As shown in Table 1, the use of facial expressions significantly (t-test with $p < 0.001$) outperforms the random baseline in each condition. The highest accuracy is achieved for the *competitive facial expression* condition, followed by the *competitive* condition. This can be explained by the increased facial expressivity due to the competitive setting and posed facial expressions. As expected, the proposed method provides higher accuracies for facial expression conditions.

3.2 Learning from Facial Feedback

We compare the average learning performance of the four conditions in terms of learning from keypress feedback, learning from binary keypress feedback (equivalent to 100% accurate prediction), learning from random feedback (50% prediction accuracy) and learning from predicted binary feedback, as shown in Figure 1. Our model with 62%-79% prediction accuracy for these four conditions are in the middle of learning from binary keypress and random feedback. Our experimental results in Figure 1 show that, when the prediction accuracy is low in the first three conditions (control, facial expression and competitive condition), agent learning from predicted binary feedback is only a little better than learning from random feedback. However, when the prediction accuracy increased to 79% in the competitive facial expression condition, agent learning from predicted binary feedback with our model can reach to around 10 which is close to the performance of learning from binary keypress feedback (around 20). Therefore, this suggest that learning solely from predicted feedback based on facial expressions is possible and there is still much room for improvement in agent’s performance using improved models with higher prediction accuracy.

ACKNOWLEDGMENTS

This work was partially supported by Natural Science Foundation of China (under grant No. 51809246), Natural Science Foundation of Shandong Province (under grant No. ZR2018QF003). This research was part of Science Live, the innovative research program of Science Center NEMO in Amsterdam that enables scientists to carry out real, publishable, peer-reviewed research using NEMO visitors as volunteers. Science Live is partially funded by KNAW and NWO.

REFERENCES

[1] Dilip Arumugam, Jun Ki Lee, Sophie Saskin, and Michael L Littman. 2019. Deep reinforcement learning from policy-dependent human feedback. *arXiv preprint*

arXiv:1902.04257 (2019).

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

[3] W Bradley Knox, Brian D Glass, Bradley C Love, W Todd Maddox, and Peter Stone. 2012. How humans teach agents. *International Journal of Social Robotics* 4, 4 (2012), 409–421.

[4] W Bradley Knox and Peter Stone. 2009. Interactively shaping agents via human reinforcement: The TAMER framework. In *Proceedings of the fifth international conference on Knowledge capture*. ACM, 9–16.

[5] Guangliang Li, Randy Gomez, Keisuke Nakamura, and Bo He. 2019. Human-Centered Reinforcement Learning: A Survey. *IEEE Transactions on Human-Machine Systems* 49, 4 (2019), 337–349.

[6] Guangliang Li, Hayley Hung, Shimon Whiteson, and W Bradley Knox. 2013. Using informative behavior to increase engagement in the tamer framework. In *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*. 909–916.

[7] Guangliang Li, Hayley Hung, Shimon Whiteson, and W Bradley Knox. 2014. Learning from human reward benefits from socio-competitive feedback. In *4th Joint IEEE International Conferences on Development and Learning and Epigenetic Robotics,*. IEEE, 93–100.

[8] Guangliang Li, Shimon Whiteson, W. Bradley Knox, and Hayley Hung. 2016. Using informative behavior to increase engagement while learning from human reward. *Autonomous Agents and Multi-Agent Systems* 30, 5 (2016), 826–848.

[9] Robert Loftin, Bei Peng, James MacGlashan, Michael L Littman, Matthew E Taylor, Jeff Huang, and David L Roberts. 2015. Learning behaviors via human-delivered discrete feedback: modeling implicit feedback strategies to speed up learning. *Autonomous Agents and Multi-Agent Systems* (2015), 1–30.

[10] Patrick M Pilarski, Michael R Dawson, Thomas Degris, Farbod Fahimi, Jason P Carey, and Richard S Sutton. 2011. Online human training of a myoelectric prosthesis controller via actor-critic reinforcement learning. In *Rehabilitation Robotics (ICORR), 2011 IEEE International Conference on*. IEEE, 1–7.

[11] Andrea L Thomaz and Cynthia Breazeal. 2008. Teachable robots: Understanding human teaching behavior to build more effective robot learners. *Artificial Intelligence* 172, 6 (2008), 716–737.

[12] Priscilla L Vail. 1994. *Emotion: The on/off switch for learning*. Modern Learning Press.

[13] Garrett Warnell, Nicholas Waytowich, Vernon Lawhern, and Peter Stone. 2018. Deep tamer: Interactive agent shaping in high-dimensional state spaces. In *Thirty-Second AAAI Conference on Artificial Intelligence*.