Towards Learning from Implicit Human Reward

(Extended Abstract)

Guangliang Li[*], Hamdi Dibeklioğlu[†], Shimon Whiteson[‡] and Hayley Hung[†]
[*]Ocean University of China, Qingdao, China & University of Amsterdam, Amsterdam, The Netherlands
[†]Delft University of Technology, Delft, The Netherlands
[‡]University of Oxford, Oxford, UK
g.li@uva.nl, {h.dibeklioglu, h.hung}@tudelft.nl, shimon.whiteson@cs.ox.ac.uk

ABSTRACT

The *TAMER* framework provides a way for agents to learn to solve tasks using human-generated rewards. Previous research showed that humans give copious feedback early in training but very sparsely thereafter and that an agent's competitive feedback — informing the trainer about its performance relative to other trainers — can greatly affect the trainer's engagement and the agent's learning. In this paper, we present the first large-scale study of TAMER, involving 561 subjects, which investigates the effect of the agent's competitive feedback in a new setting as well as the potential for learning from trainers' facial expressions. Our results show for the first time that a TAMER agent can successfully learn to play Infinite Mario, a challenging reinforcement-learning benchmark problem. In addition, our study supports prior results demonstrating the importance of bi-directional feedback and competitive elements in the training interface. Finally, our results shed light on the potential for using trainers' facial expressions as reward signals, as well as the role of age and gender in trainer behavior and agent performance.

Categories and Subject Descriptors

I 2.6 [**Artificial Intelligence**]: Learning

General Terms

Performance, Human Factors, Experimentation

Keywords

Reinforcement learning; human agent interaction

1. INTRODUCTION

Socially intelligent autonomous agents have the potential to become our high-tech companions in the family of the future. 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.

Learning from human reward, i.e., evaluations of the quality of the agent's behavior, has proven to be a powerful technique for facilitating the teaching of artificial agents by their human users [2, 8, 4]. Compared to learning from demonstration [1], learning from human reward does not require the human to be able to perform the task well herself; she needs only to be a good judge of performance. Nonetheless, agent learning from human reward is limited by the quality of the interaction between the human trainer and the agent.

Previous research shows that the interaction between the agent and the trainer should ideally be bi-directional [5, 6, 7] and that if an agent informs the trainer about the agent's past and current performance and its performance relative to others, the trainer will provide more feedback and the agent will ultimately perform better. This paper presents the results of the first large-scale study of TAMER—a popular method for enabling autonomous agents to learn from human reward [4]—by implementing it in the Infinite Mario domain. Our study was conducted at a science museum in Amsterdam using 561 museum visitors as subjects and investigates the effect of the agent's socio-competitive feedback in a new setting. In addition, we also study the potential of using facial expressions as reward signals, since several TAMER studies have shown that humans give copious feedback early in training but very sparsely thereafter [3, 5].

Our experimental results show for the first time that a TAMER agent can successfully learn to play Infinite Mario, a challenging reinforcement learning benchmark problem. Moreover, our study provides large-scale support of the results of Li et al. [5, 6] demonstrating the importance of bi-directional feedback and competitive elements in the training interface and sheds light on the potential for using trainers' facial expressions as reward signals, as well as the role of age and gender in trainer behavior and agent performance.

2. EXPERIMENT CONDITIONS

In our study at the science museum in Amsterdam involving 561 subjects, we test two independent variables: 'competition'—whether the agent will inform the competitive feedback to the trainer, and 'facial expression'—whether trainers were told that their facial expressions would be used in addition to key presses to train the agent. The main idea of the facial expression condition is to examine the effect that the additional modality of facial expressions could have on the cognitive load of trainers and whether this varies depending on age or gender.

We investigate how 'competition' and 'facial expression' affect the agent's learning performance and trainer's facial expressiveness in four experimental conditions in our study: the *control condition*—without 'competition' or 'facial expression', the *facial expression condition*—without 'compe-

Appears in: *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2016), J. Thangarajah, K. Tuyls, S. Marsella, C. Jonker (eds.), May 9–13, 2016, Singapore.*

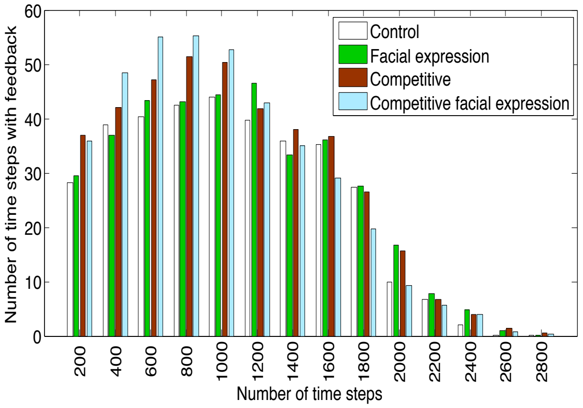

Figure 1: Mean number of time steps with feedback per 200 time steps for all four conditions during the training process.

tition' 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. In addition, we expect that both 'competition' and 'facial expression' will increase the trainer's facial expressiveness.

3. EXPERIMENTAL RESULTS

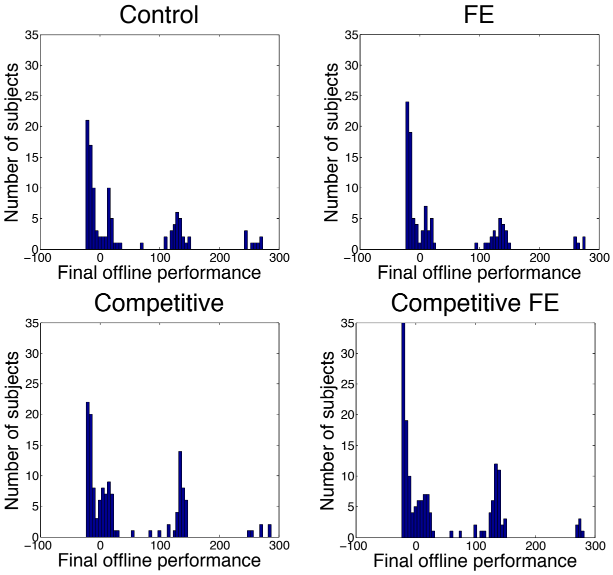

Figure 2: Distribution of final offline performance across the four conditions. FE=Facial Expression.

Figure 1 shows how feedback was distributed per 200 time steps over the learning process for the four conditions. From Figure 1 we can see that the number of time steps with feedback received by agents in the four conditions increased at the early training stage and decreased dramatically afterwards, which supports previous studies [3, 5] and our motivation for investigating methods of enabling agents to learn from the trainer's facial expressions. In addition, it shows that the agent's competitive feedback can increase the number of feedback given by the trainer before 1000 time steps. Figure 2 shows histograms of the distribution of the final offline performance for the four conditions. Further analysis with n-way ANOVA shows that 'competition' can significantly improve agent learning ($p = 0.035$) and help the best trainers the most ($p = 0.01$). In addition, our results suggest that 'facial expression' has a significantly negative effect on agent training by female subjects, especially those who are less than 13 years old ($p = 0.008$) and cannot train agents to perform well ($p = 0.01$).

Furthermore, our analysis shows that telling trainers to use facial expressions makes them inclined to exaggerate their expressions, resulting in higher accuracy for predicting positive and negative feedback using facial expressions. Competitive conditions also elevated facial expressiveness and further increased predicted accuracy. This has significant consequences for the design of agent learning systems that wish to take into account a trainer's spontaneous facial expressions as a reward signal. Further investigation into the nature of spontaneous and posed facial expressions is needed, in particular in terms of their relation to feedback quality and quantity.

Acknowledgments

This research was part of Science Live, the innovative research programme 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] B. D. Argall, S. Chernova, M. Veloso, and B. Browning. A survey of robot learning from demonstration. *Robotics and autonomous systems*, 57(5):469–483, 2009.

[2] C. Isbell, C. R. Shelton, M. Kearns, S. Singh, and P. Stone. A social reinforcement learning agent. In *Proceedings of the fifth international conference on Autonomous agents*, pages 377–384. ACM, 2001.

[3] W. B. Knox, B. D. Glass, B. C. Love, W. T. Maddox, and P. Stone. How humans teach agents. *International Journal of Social Robotics*, 4(4):409–421, 2012.

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

[5] G. Li, H. Hung, S. Whiteson, and W. B. Knox. Using informative behavior to increase engagement in the TAMER framework. In *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*, pages 909–916, 2013.

[6] G. Li, H. Hung, S. Whiteson, and W. B. Knox. Learning from human reward benefits from socio-competitive feedback. In *Proceedings of the Fourth Joint IEEE International Conference on Development and Learning and on Epigenetic Robotics*, pages 93–100, 2014.

[7] G. Li, S. Whiteson, W. B. Knox, and H. Hung. Using informative behavior to increase engagement while learning from human reward. *Autonomous Agents and Multi-Agent Systems*, pages 1–23, 2015.

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