Socially Intelligent Autonomous Agents that Learn from Human Reward

G. Li

In the future, autonomous agents will operate in human inhabited environments in many real world applications and become an integral part of human's daily lives. Therefore, when autonomous agents enter into the real world, they need to adapt to many novel, dynamic and complex situations that cannot be imagined and preprogrammed in the lab and try to learn new skills. Meanwhile, human users may also want to teach agents behaviors they like. Interactive shaping is a teaching method that has been developed and proven to be a powerful technique for enabling autonomous agents to learn how to perform tasks according to a human trainer's preference. In interactive shaping, an agent learns from real time human reward signals, i.e., evaluations of the quality of an agent's behavior delivered by a human observer. However, though the agent can already learn from such human-delivered reward signals, the agent learning critically depends on the quality and quantity of the interaction between the human teacher and the agent. In this thesis, we use interactive shaping, the TAMER framework in particular, to investigate methods for increasing the efficiency of agent learning from human reward. We consider solutions to better engage a human trainer from the perspectives of both the agent and human trainer.

First, we study what information the agent should share with the human trainer by proposing a bi-directional approach. We implement our approach by facilitating the agent to provide two kinds of informative feedback: uncertainty and performance, in addition to allowing a human trainer to give the agent rewards. The empirical results of our approach with a 51-user study in the game of Tetris show that the agent's informative feedback can significantly increase the duration of training and the amount of feedback provided by the trainer, with the uncertainty-informative interface generating the most feedback. As we expected, our results showed that the performance-informative feedback led to substantially better agent performance. However, the uncertainty-informative feedback led to worse agent performance although much more human feedback was elicited. The results that measuring performance increased performance, and measuring uncertainty reduced uncertainty, also fitted with a pattern observed in organizational behavior research that follows from the adage "you get what you measure" — sharing behavior-related metrics will tend to make people attempt to improve their scores with respect to those metrics. This provides a potentially powerful guiding principle for human-agent interface design in general.

Next, we seek to build on our approach by investigating how competition not only among strangers but also among close friends or family members in remote as well as co-located scenarios impacts agent training by human trainers and agent final learning performance in two user studies. In the first study with 85 subjects via a Facebook app, we test the effect of the passive and active socio-competitive feedback — agent's performance relative to other trainers — in the game of Tetris. The results of our user study show that the agent's social feedback can induce the

trainer to train statistically significant longer and give much more feedback and the agent performance is much better when social-competitive feedback is provided. The second user study was conducted at NEMO (the Dutch national science museum) with 561 subjects (209 adults and 352 children). In this study, we further investigate the effect of socio-competitive feedback among closely related trainers with a much broader range of ages and genders who trained agents at the same time in the same room. The results show for the first time that a TAMER agent can successfully learn to play Infinite Mario, a challenging reinforcement learning benchmark problem based on the popular video game. Our study also provides large-scale support of the results of prior studies demonstrating the importance of bi-directional feedback and competitive elements in the training interface.

Lastly, from the human's point of view, as time progresses, humans may get tired of giving explicit feedback (e.g., button presses to indicate positive or negative reward). Therefore, in our NEMO study, we investigate the potential of using facial expressions as reward signals, which have often been used by humans to consciously or subconsciously encourage or discourage specific behaviors they want to teach. Our results show the statistically significant negative effect of telling female trainers to use facial expressions as a separate channel to train agents. Moreover, our analysis shows that when using facial expressions to classify positive and negative keypress feedback, the use of facial expressions significantly outperforms the random baseline, and 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. Furthermore, competition can elevate facial expressiveness and further increase the predicted accuracy. These results suggest that while the use of facial expressions for agent training seems sensible, there are a number of factors to consider relating to additional cognitive load of making posed facial expressions which leads to a reduced quality in keypress feedback.

The work in this thesis has made several contributions towards the understanding of interactive shaping by developing methods to augment the agent's learning through it and further ease the human teacher's cognitive load. Our results may generalize to other task domains and apply to other interactive learning algorithms.