TAMER


Trick
Shaping
McCANN DOGS

Interactively Shaping Agents via Human Reinforcement

The TAMER Framework

W. Bradley Knox
Department of Computer Science
The University of Texas at Austin
bradknox@cs.utexas.edu

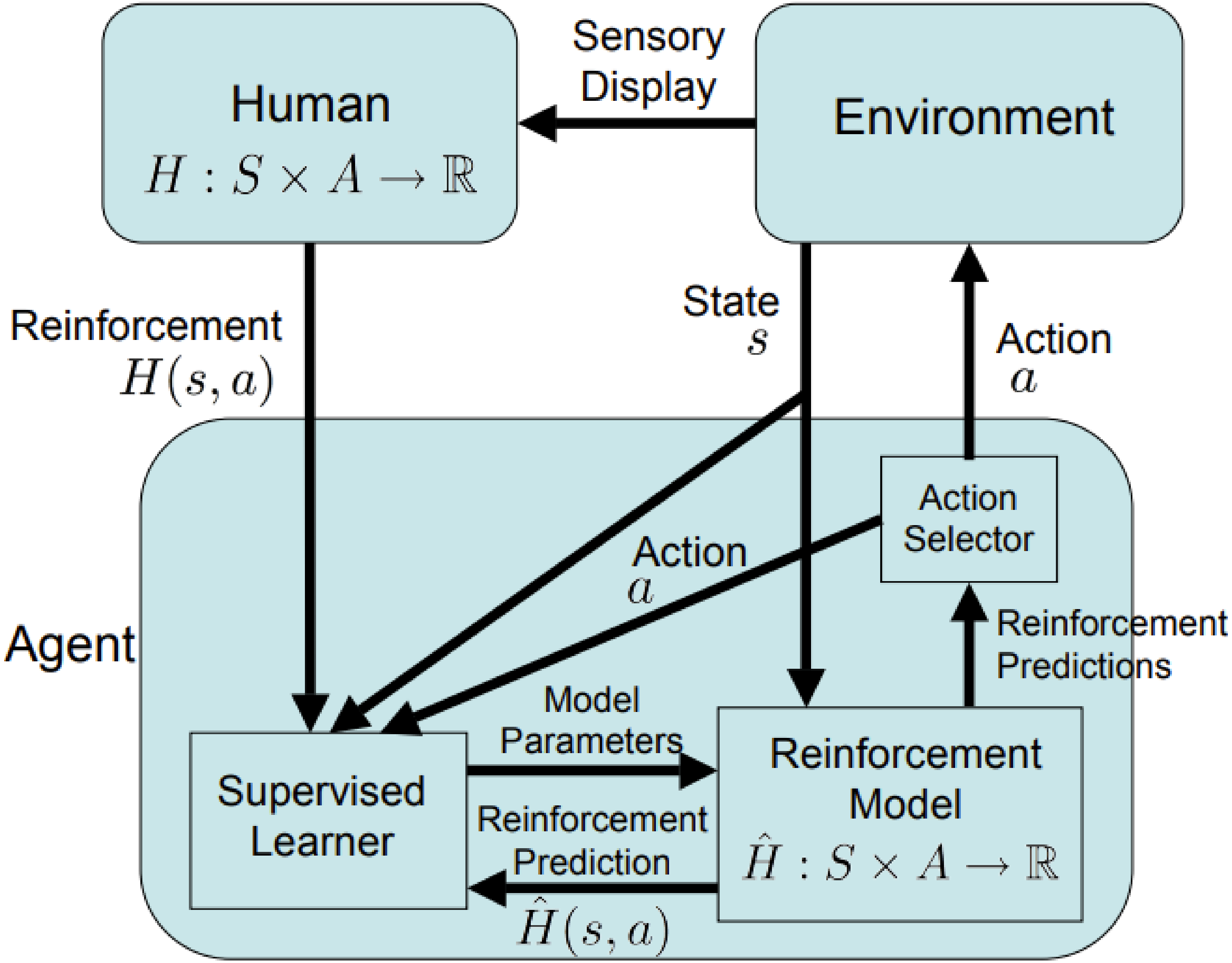
Peter Stone
Department of Computer Science
The University of Texas at Austin
pstone@cs.utexas.edu

ABSTRACT

As computational learning agents move into domains that incur real costs (e.g., autonomous driving or financial investment), it will be necessary to learn good policies without numerous high-cost learning trials. One promising approach to reducing sample complexity of learning a task is knowledge transfer from humans to agents. Ideally, methods of transfer should be accessible to anyone with task knowledge, regardless of that person's expertise in programming and AI. This paper focuses on allowing a human trainer to interactively shape an agent's policy via reinforcement signals. Specifically, the paper introduces "Training an Agent Manually via Evaluative Reinforcement," or TAMER, a framework that enables such shaping. Differing from previous

deploy these agents in real-world domains, making decisions that affect our lives. However, with real-world deployment comes real-world costs. For such a deployment to be viable, agents will not be able to use hundreds or thousands of learning trials to reach a good policy when each suboptimal trial is costly. For example, an autonomous driving agent should not learn to drive by crashing into road barriers and endangering the lives of pedestrians.

Fortunately, for many of these tasks, humans have domain knowledge that could speed the learning process, reducing costly sample complexity. Currently, most knowledge transfer from humans to agents occurs via programming, which is time-consuming and inaccessi-


Human
$H : S \times A \rightarrow \mathbb{R}$
Sensory Display
Environment
Reinforcement $H(s,a)$
State s
Action a
Agent
Action Selector
Action a
Reinforcement Predictions
Model Parameters
Supervised Learner
Reinforcement Prediction $\hat{H}(s,a)$
Reinforcement Model
$\hat{H} : S \times A \rightarrow \mathbb{R}$

Deep TAMER: Interactive Agent Shaping in High-Dimensional State Spaces

Garrett Warnell[1], Nicholas Waytowich[1,2], Vernon Lawhern[1], Peter Stone[3]
[1]U.S. Army Research Laboratory, [2]Columbia University, New York, [3]The University of Texas at Austin
{garrett.a.warnell.civ,nicholas.r.waytowich.civ}@mail.mil,
vernon.j.lawhern.civ@mail.mil,pstone@cs.utexas.edu

Abstract

While recent advances in deep reinforcement learning have allowed autonomous learning agents to succeed at a variety of complex tasks, existing algorithms generally require a lot of training data. One way to increase the speed at which agents are able to learn to perform tasks is by leveraging the input of human trainers. Although such input can take many forms, real-time, scalar-valued feedback is especially useful in situations where it proves difficult or impossible for humans to provide expert demonstrations. Previous approaches have shown the usefulness of human input provided in this fashion (e.g., the TAMER framework), but they have thus far not considered high-dimensional state spaces or employed the use of deep learning. In this paper, we do both: we propose *Deep TAMER*, an extension of the TAMER framework that lever-

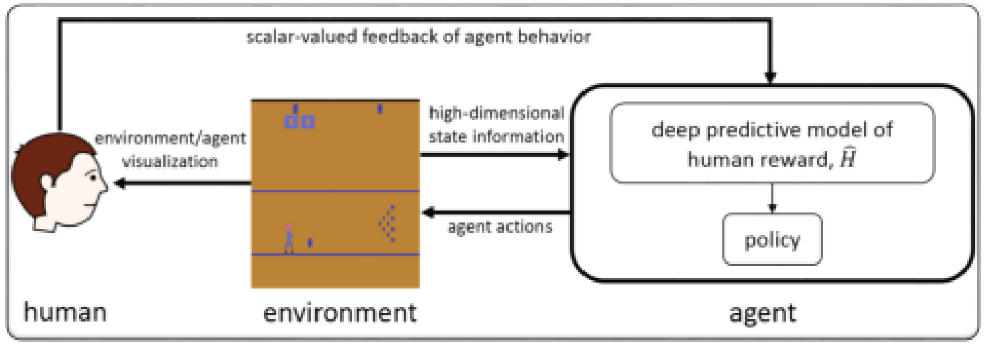


Figure 1: The Deep TAMER framework proposed in this paper. A human observes an autonomous agent trying to perform a task in a high-dimensional environment and provides scalar-valued feedback as a means by which to shape agent behavior. Through the interaction with the human, the agent learns the parameters of a deep neural network, $\hat{H}$, that is used to predict the human's feed-

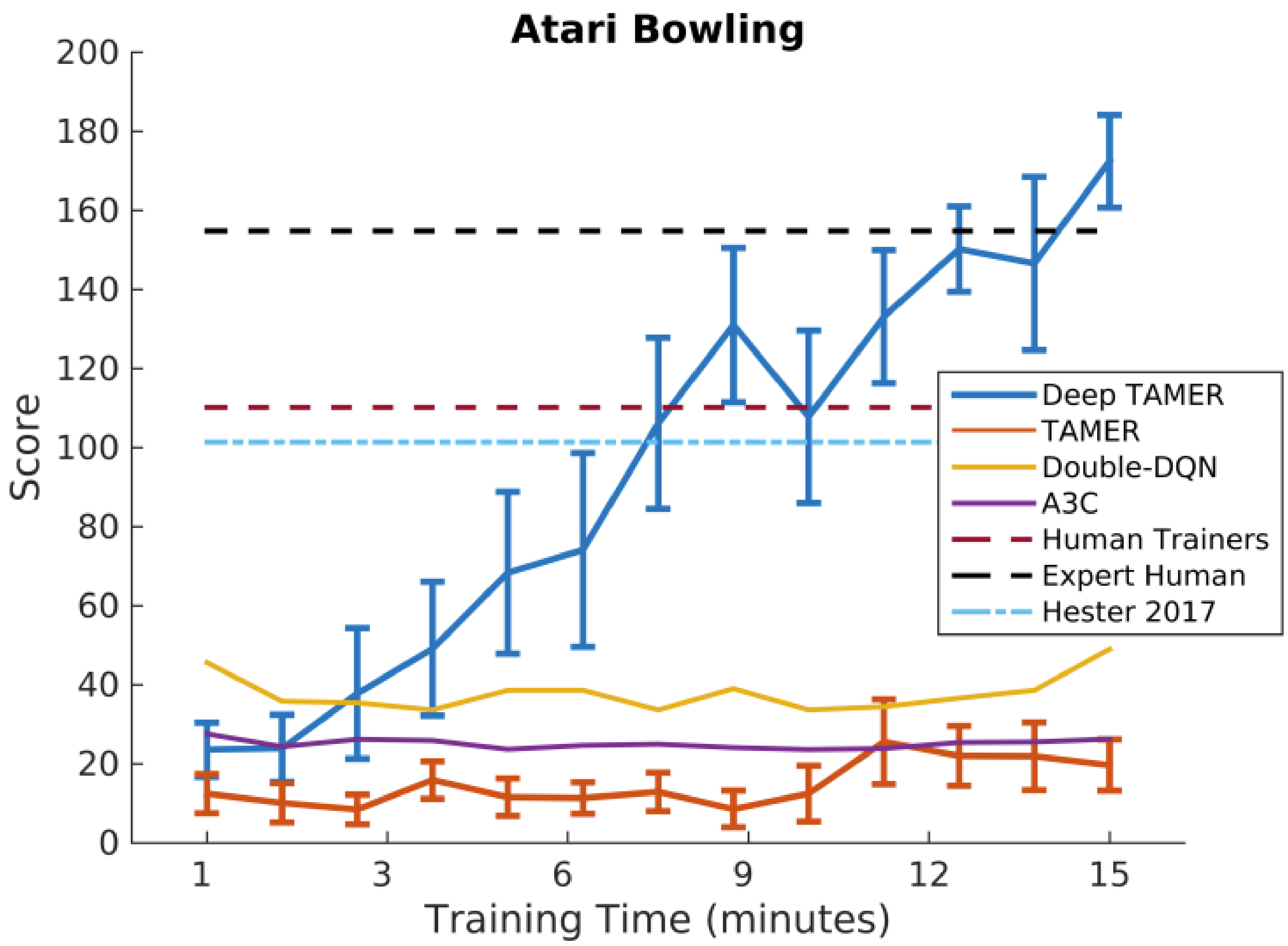
Double DQN

- Q-Learning update

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \alpha(Y_t^{\mathrm{Q}} - Q(S_t, A_t; \boldsymbol{\theta}_t))\nabla_{\boldsymbol{\theta}_t} Q(S_t, A_t; \boldsymbol{\theta}_t)$$

$$Y_t^{\mathrm{DQN}} \equiv R_{t+1} + \gamma \max_a Q(S_{t+1}, a; \boldsymbol{\theta}_t^-)$$

$$Y_t^{\mathrm{DoubleDQN}} \equiv R_{t+1} + \gamma Q(S_{t+1}, \operatorname*{argmax}_a Q(S_{t+1}, a; \boldsymbol{\theta}_t), \boldsymbol{\theta}_t^-)$$


Atari Bowling
Score
Training Time (minutes)
0
20
40
60
80
100
120
140
160
180
200
1
3
6
9
12
15
Deep TAMER
TAMER
Double-DQN
A3C
Human Trainers
Expert Human
Hester 2017