Social Robotics - TP2 Learning From Human Feedback

Instructors: Mohamed Chetouani, Silvia Tulli, Louis Simon

In this lab session, you will learn more about learning from feedback, where humans provide evaluative feedback as a form of reward to guide the learning agent's behavior. We will focus on TAMER, a well-known interactive reinforcement learning algorithm in which the agent is trained manually using human feedback. In TAMER, the standard reward function $R(s, a)$ is replaced by a human reward $\hat{H}_\theta(s, a)$, which is estimated from human feedback h. Learning consists of fitting a regression model to approximate the human reward function.

Question 1

To familiarize yourself with OpenAI Gym, explore other OpenAI Gym environments and apply the TAMER algorithm in a different environment. Continue using the Pygame structure. Make sure you understand the observation space, action space, rewards, and termination conditions. `https://gymnasium.farama.org/`

Question 2

Suggest alternative ways to incorporate human feedback by examining different variants of the TAMER algorithm. Try implementing one variant of your choice.

Question 3

In this practical session, you used the keyboard for feedback. List alternative interfaces for collecting human feedback, and discuss the advantages and disadvantages of each.

Question 4

What are the three fundamental differences between a reward function and a human feedback function? Specify the difference in the learning update rule between the two.

References

1. Knox, W.B., & Stone, P. (2009). *Interactively shaping agents via human reinforcement: the TAMER framework*. International Conference on Knowledge Capture. `https://www.cs.utah.edu/~dsbrown/readings/tamer.pdf`

2. Knox, W.B., & Stone, P. (2012). *Reinforcement learning from simultaneous human and MDP reward*. Adaptive Agents and Multi-Agent Systems. `https://www.cs.cmu.edu/~jeanoh/16-785/papers/knox-aamas2012-tamer.pdf`

3. Warnell, G., Waytowich, N.R., Lawhern, V.J., & Stone, P. (2017). *Deep TAMER: Interactive Agent Shaping in High-Dimensional State Spaces*. ArXiv, abs/1709.10163. `https://arxiv.org/pdf/1709.10163.pdf`

4. Arakawa, R., Kobayashi, S., Unno, Y., Tsuboi, Y., & Maeda, S. (2018). *DQN-TAMER: Human-in-the-Loop Reinforcement Learning with Intractable Feedback*. ArXiv, abs/1810.11748. `https://www.semanticscholar.org/reader/31ed72a18ed8a1008786130af9f1d61761cff4f3`

5. Karalus, J., & Lindner, F. (2021). *Accelerating the Learning of TAMER with Counterfactual Explanations*. 2022 IEEE International Conference on Development and Learning (ICDL), 362–368. `https://www.semanticscholar.org/reader/243abd03e6fa267219d5afd82f387dddc0db26a3`