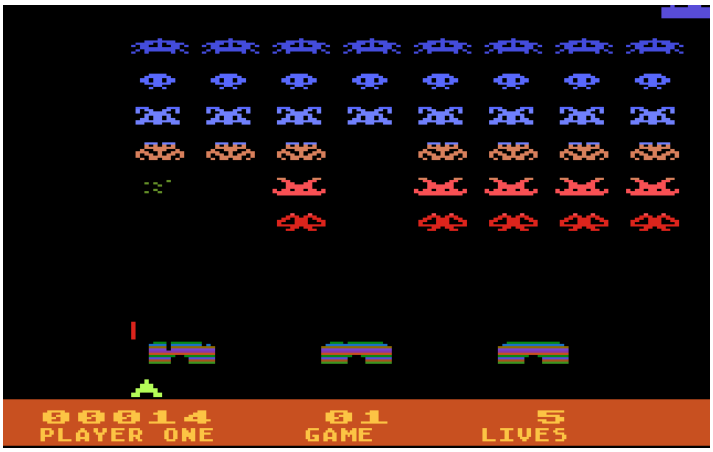
AIND

RESEARCH REVIEW

GIL AKOS // 17 JUNE 2017

DEEP REINFORCEMENT LEARNING FROM HUMAN PREFERENCES

Christiano et al., 2017 // arXiv:1706.03741 [stat.ML]



https://r.mprd.se/Atari%205200/Screenshots/S/Space_Invaders_1.png

Many games that we understand to be relatively basic are actually characterized by a degree of complexity that challenges the creation and definition of Artificially Intelligent agents. In contrast to Isolation, Chess, or even Go, in which actions update the environment across a single time-step, the reward given for actions in early Atari games is dependent upon actions that play out over time or are contingent upon a sequence of actions. This condition makes well-defined reward functions difficult to explicitly define as well as limits our ability as humans to create direct demonstrations of what should be understood as successful behavior. In “Deep Reinforcement Learning from Human Preferences” by Christiano et al., 2017, the authors seek to define an effective and efficient system for solving complex reinforcement learning tasks, like Space Invaders, without access to a reward function but instead by having humans provide feedback on short videos of agents in action called trajectory segments.

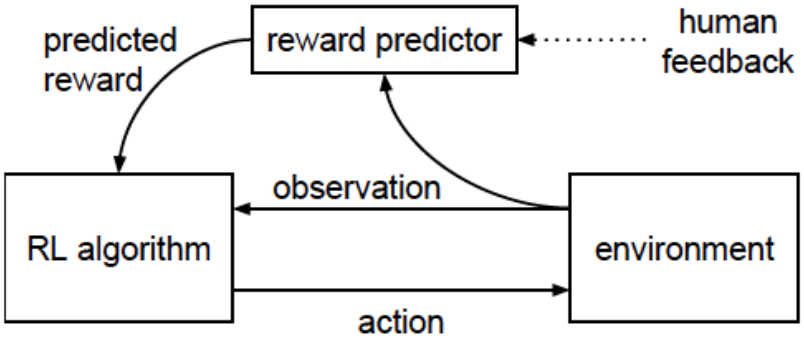
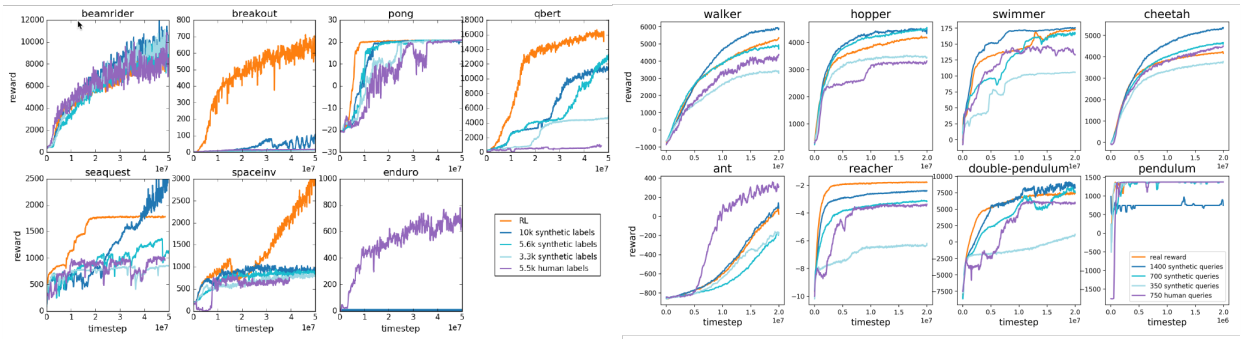


Figure 1 from arXiv:1706.03741 [stat.ML]

This research is compelling for its problem-solving architecture, the scale of the results, and its future impact for ongoing research in AI. First, and most likely due to the inability to define a reward function for the games of choice, the study employs a nested machine learning architecture. In Isolation and other board games, the environment is static and we can define a range of reward functions that directly correlate to the result of a forecasted move and deploy supervised learning techniques. In the Atari games studied, this direct correlation is not present because of the dynamic nature of the environment therefore reward functions are difficult to create. This study utilizes supervised learning to help create the reward function itself with the aid of non-expert trainers, which is then incorporated into a reinforcement learning model. This nested approach to the techniques unlocks the ability of the overall system to learn complex behaviors and demonstrates significant results both in terms of game-playing performance and efficiency. Second, in terms of the scale of results, human oversight was provided for only 1% of the interactions between the agent and the environment, yet across the games played the results show substantial learning on most and even exceeding standard reinforcement learning techniques on some games. This range of results is due to the differing dynamics at play in each game and the ability of the human observers to understand and label the video segments accurately. Furthermore, the study shows positive results for complex tasks can be achieved economically. The architecture used paired with minimal human oversight allows the agent-environment interactions to do the heavy lifting in terms of learning, thereby "reducing the interaction complexity by roughly 3 orders of magnitude." To demonstrate the flexibility of the approach, the authors also applied the architecture to simulated robotics with similar productive results.



Figures 3, 2 from arXiv:1706.03741 [stat.ML]

Lastly, this research expands the range of new problems that can be addressed with Artificial Intelligence. Problems thought too complex now have a reliable method for experimentation, and one that can be done with reasonable cost and today's technological capabilities. Additionally, the approach as tested through both Atari games and robotics shows promise for generalizable use across problem domains. Beyond games this research "represents a step towards practical applications of deep RL to complex real-world tasks."