A comparative study of Reinforcement Learning techniques over multiple OpenAI Gym environments

Anubhav Shrimal, Mahika Wason, Rupal Jain
Advisors: Dr. Mayank Vatsa, Prabhat Kumar

Artificial Intelligence CSE643

ABSTRACT

In this project, we try solving real-world problems' virtual simulations in Open AI gym using the reinforcement paradigm of an actor-critic. An actor performs specific actions in its environment which are rewarded by the critic. The actor learns from these feedbacks and carries out future activities with an aim to improve upon them. We analyze and compare the performance of different reinforcement based techniques over multiple OpenAI Gym environments.

MOTIVATION AND BACKGROUND

- Reinforcement learning is a reward based machine learning technique. The associated agent executes actions with an aim to maximise its cumulative reward. On contrary to other ML techniques where the agent knows about its environment apriori, RL agents adapt to their environment over time and discover the most rewarding actions with trial and error.
- RL is generally used for solving gaming-based problems like Breakout and Pong to name a few. RL is also widely used for teaching human-like limbs' movements to virtual humanoids (and other human-like creatures). Such techniques have direct applications in real world scenarios such as training prosthetics to act like natural limbs for the handicapped. These applications motivate us to compare and contrast between different RL based approaches and thus, figuring out the trade-off of using them for different situations.

ENVIRONMENTS

- We are going to use the OpenAI GYM environments for the purpose of this project.
- Open AI gym is a platform for simulating real world scenarios. These simulations enable testing of artificial intelligence algorithms and testing various different reinforcement learning algorithms.
- These environments includes Atari games, humanoid robots and other complex animal simulations.

OBJECTIVES

- Our objective is to implement and run a set of RL based algorithms on different virtual environments.
- The environments may vary from being distinct or continuous, dynamic or static, partially or fully observable to name a few.
- RL based algorithms vary from Monte Carlo to Q learning and other more advanced Deep RL techniques.
- Based on the inferences, we aim to reason out why one algorithm works better than the other on a particular environment, and why the same algorithm works better on another.

Image source: Google Search

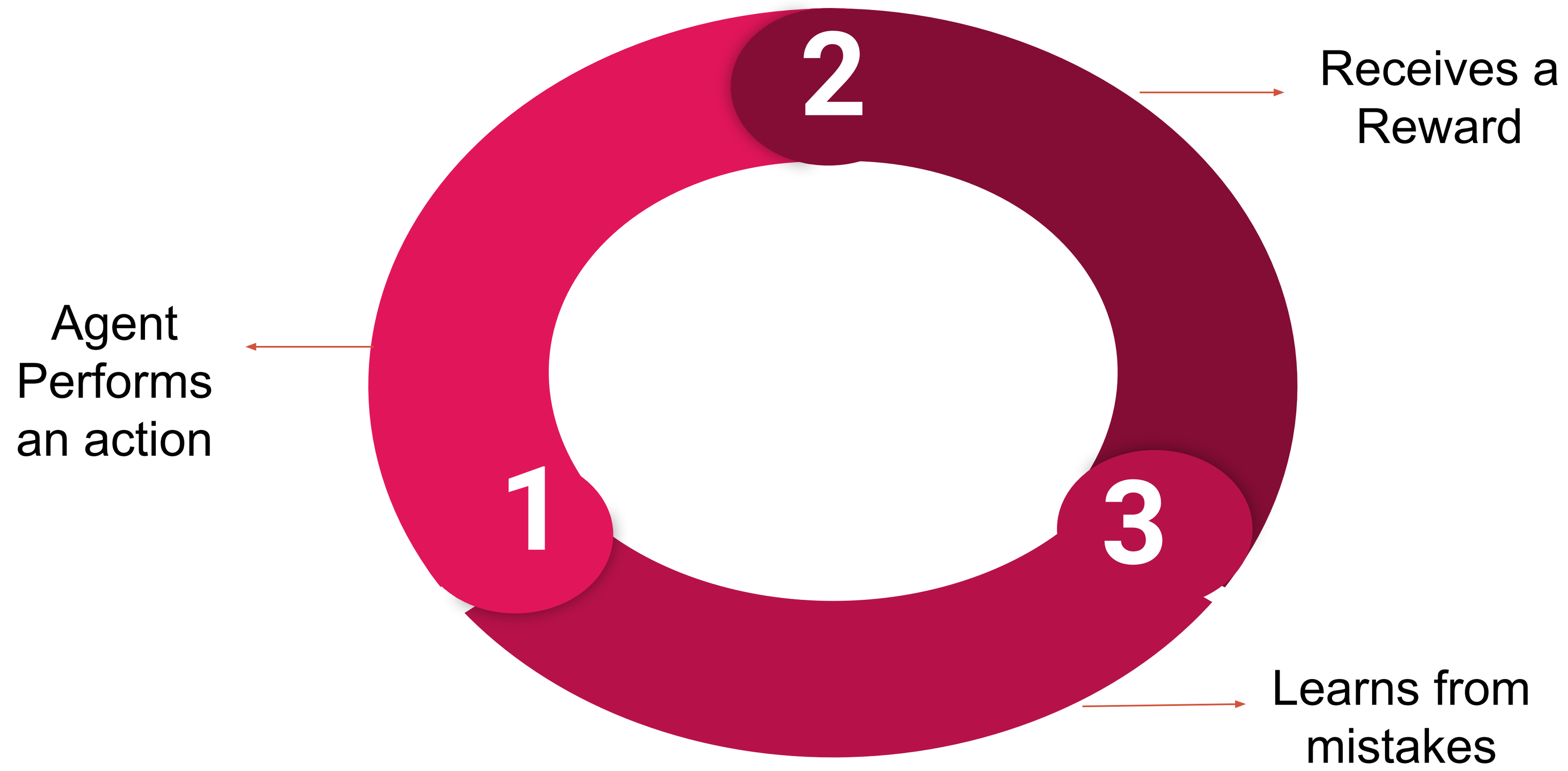

Fig. Reinforcement Learning Paradigm

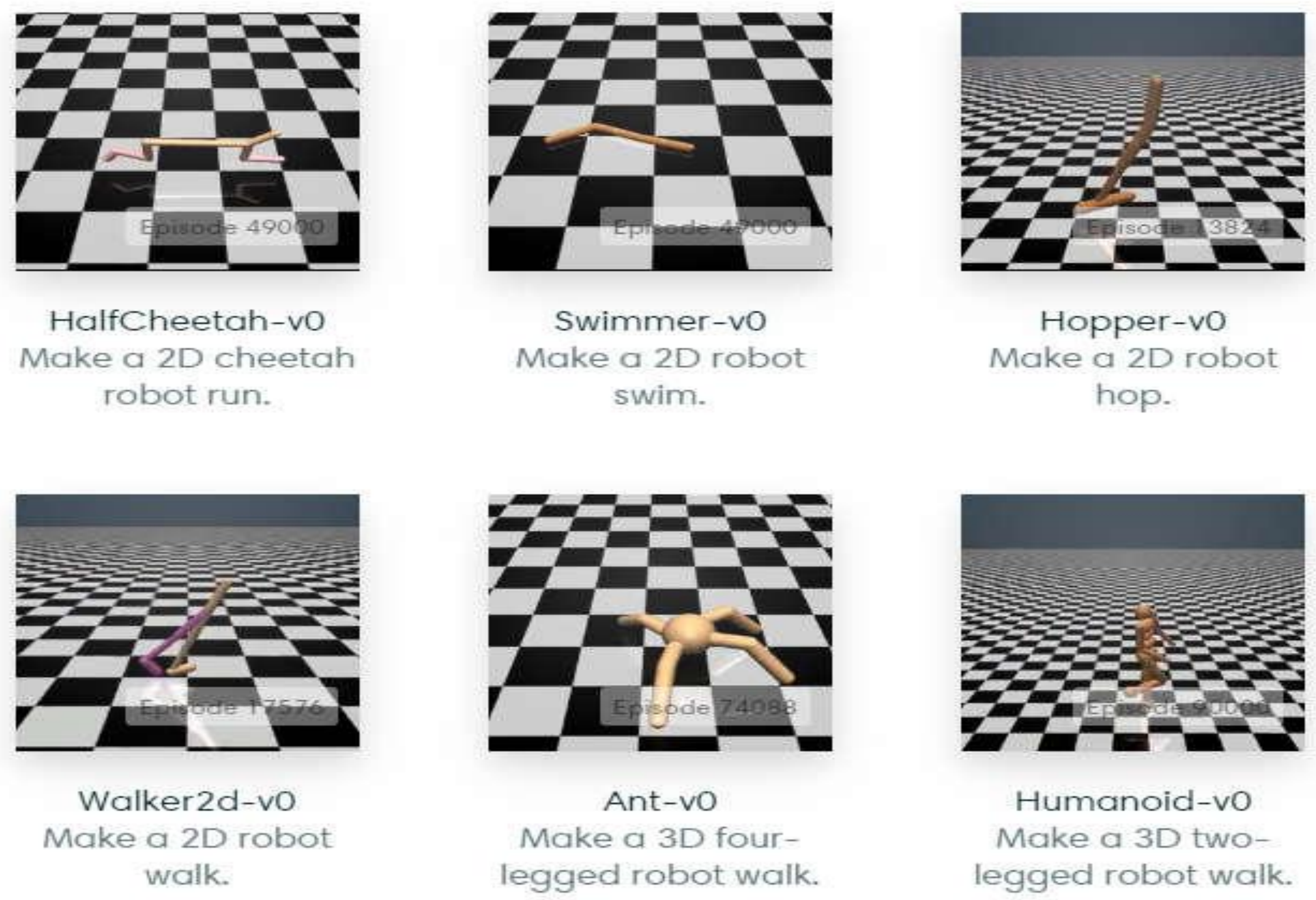

Image source: Google Search

EVALUATION METRIC

- RL algorithms follow a reward based learning approach. Hence the evaluation metrics applied in supervised learning based approaches or other paradigms cannot be applied.
- The evaluation of our algorithms would be how much reward the agent achieves in an environment.
- The rewards are specific to an environment. For example, in atari games it would be to achieve higher score whereas in humanoid robots it would be to walk as far as possible.
- These rewards are provided by the gym environment itself and would help us form our evaluation metric.

ANALYSIS

- Each algorithm's performance will be compared on the basis of:
  - Time taken to train the model till a certain reward is achieved.
  - Space/Memory consumption.
- We may also change various parameters such as value of the reward function and other hyper parameters and see how they perform.
- Various other analysis factors will be incorporated as we progress in the project.

MILESTONES

- By the first evaluation, we aim to implement Monte Carlo and Q-Learning technique and simulate it over different OpenAI Gym environments. And then, compare its performance over those environments.
- By the second evaluation, we aim to implement other RL algorithms on the same set of environments. And finally, perform a comparative study based on the results obtained from all the chosen algorithms.

REFERENCES

1. Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M. (2013). Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*.
2. Mataric, M. (1991). A comparative analysis of reinforcement learning methods.
3. Morimoto, J., Cheng, G., Atkeson, C. G., & Zeglin, G. (2004, April). A simple reinforcement learning algorithm for biped walking. In Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on (Vol. 3, pp. 3030-3035). IEEE.
4. Kirchner, F. (1998). Q-learning of complex behaviours on a six-legged walking machine. *Robotics and Autonomous Systems*, *25*(3-4), 253-262.
5. Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., ... & Wierstra, D. (2015). Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*.
6. Heess, N., Sriram, S., Lemmon, J., Merel, J., Wayne, G., Tassa, Y., ... & Silver, D. (2017). Emergence of locomotion behaviours in rich environments. *arXiv preprint arXiv:1707.02286*.
7. Mohan, V. (2014). A comparison of various reinforcement learning algorithms to solve racetrack problem.
8. Ma, B., Tang, M., & Zhang, J. Exploration of Reinforcement Learning to SNAKE.