Introduction

Reinforcement learning is a machine learning training method where desired behavior is rewarded while undesirable behavior is penalized. These are called positive and negative reinforcement respectively. A reinforcement learning agent is capable of observing and interpreting its environment, acting then receiving a reward or penalty based on its action. A reinforcement learning agent is said to learn through trial and error.

In this paper, we will be focusing on the topic of positive reinforcement by using the game Brick Breaker as the environment. In Brick Breaker, a player bounces a ball to destroy bricks by moving a paddle left and right. The player loses if the ball leaves the screen and wins if all bricks are destroyed.

Q-Learning is the reinforcement algorithm utilized in this paper. The foundation of the code was written by Matthew Chan and is available at medium.com. The code has been modified to work with the game.

Genetic algorithm will also be employed in this paper and is used to determine the hyperparameters of the Q-Learning agent. Different techniques such as single/double crossover and tournament/roulette selection are used to ensure more variety during the evolution process to produce better results.

This paper aims to study the effects of different improvements of Q-Learning on the agent performance. Combinations of these improvements will then be tested to determine the best combination and get the best possible results. The effects of varying the game settings on the agent performance will also be studied and documented.

A GUI was also created for users to more easily change the settings of the project and streamline the process of hyperparameter tuning and data collection.

Literature Review

The main innovation in this paper is to combine existing improvements of the Q-Learning agent to produce a new, better solution. In the end, 7 research papers of interest were identified and listed below.

Firstly, **genetic algorithm** was used by Wicaksono, A. S., & Supianto, A. A. for hyperparameter optimization of the machine learning methods used in online news popularity prediction.

Secondly, H.R. Tizhoosh has introduced **opposition-based learning** as a new scheme for machine intelligence.

Thirdly, the **granularity of the state space** was shown to affect the results by Jacopo Fior and Luca Cagliero in their study of machine learning-based stock trading.

Furthermore, Michal Gregor and Juraj Spalek have also done research on optimistic exploration value functions. This has been adapted in this project as the **random initialization of the Q-tables**.

Moreover, research into action elimination with deep reinforcement learning has also been done by Tom Zahavy, et al. This has been adapted to varying the **size of the action space** in the project.

Apart from that, research into **different reward functions** on the training performance of a Double DQN has also been done by Stefan Šćepanović.

Finally, Double Q-Learning was introduced by Hado van Hasselt which lacks the overestimation bias of the Q-Learning algorithm. This has adapted to **N-tuple Q-Learning** which is another innovation in this paper where N Q-tables were created for each reinforcement learning agent.

Thus, we can study the effects of combining multiple improvements and ascertain the best combination with the best results.

Experiment

The