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.

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.

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.

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.

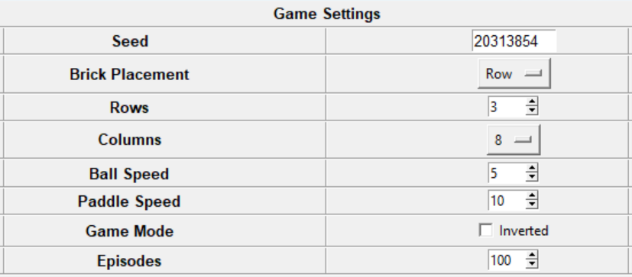
Thus, we can study the effects of combining multiple improvements and ascertain the best combination with the best results.

Graphical User Interface (GUI)

As mentioned above, users can easily change the settings of the project using the GUI created. The different settings of the GUI will be explained below.

## Game Settings

The game settings allow users to customize the different aspects of the game. The different options available to the users are shown below in Figure 1.



***Figure 1: Game Settings***

* **Seed**

|  |  |
| --- | --- |
| Range | 0 – 4294967295 |
| Default | 20313854 |
| Description | Integer used to initialize the pseudorandom number generator to ensure that results are reproducible |

* **Brick Placement**

|  |  |
| --- | --- |
| Options | Row, Column, Random |
| Default | Row |
| Description | Controls how the 3 brick types are arranged in the game. **Row** alternates by row, **Column** by column while **Random** is random. |

* **Rows**

|  |  |
| --- | --- |
| Range | 1 – 10 |
| Default | 3 |
| Description | Controls the number of brick rows in the game. |

* **Columns**

|  |  |
| --- | --- |
| Options | 1,2,3,4,5,6,8,10,12,15 |
| Default | 8 |
| Description | Controls the number of bricks in each row in the game. Can only be a factor of 600 up to 15. |

* **Ball Speed**

|  |  |
| --- | --- |
| Range | 1 – 10 |
| Default | 5 |
| Description | Controls the speed of the ball in the game. |

* **Paddle Speed**

|  |  |
| --- | --- |
| Range | 5 – 15 |
| Default | 10 |
| Description | Controls the speed of the paddle in the game. |

* **Game Mode**

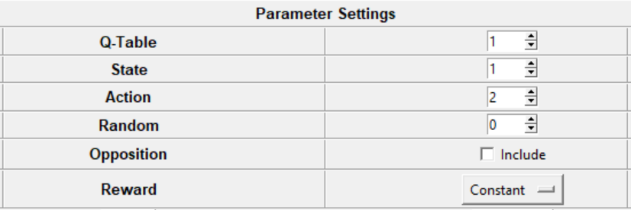
|  |  |
| --- | --- |
| Options | False, True |
| Default | False |
| Description | Allows inversion of the position of the bricks and paddle in the game. |

* **Episodes**

|  |  |
| --- | --- |
| Range | 1 – 200 |
| Default | 100 |
| Description | Controls the number of failed episodes before the game is reset and recorded as a failed run. |

## Parameter Settings

The parameter settings allow users to control the different settings of the Q-Learning agent. The different options available are shown below in Figure 2.



***Figure 2: Parameter Settings***

* **Q-Table**

|  |  |
| --- | --- |
| Range | 1 – 5 |
| Default | 1 |
| Description | Controls the number of Q-tables generated for each agent. Based on the concept of N-tuple Q-Learning which is inspired by Double Q-Learning. |

* **State**

|  |  |
| --- | --- |
| Range | 1 – 10 |
| Default | 1 |
| Description | Controls the granularity of the state space of the Q-tables. State is assigned based on distance between the ball and the paddle. State is multiplied by 2 to represent the 2 possibilities of left and right. |

* **Action**

|  |  |
| --- | --- |
| Options | 2,3 |
| Default | 2 |
| Description | Controls the size of the action space. The 2 default actions are for the paddle to move left and right. The 3rd action is to do nothing. |

* **Random**

|  |  |
| --- | --- |
| Options | None, Normal, Uniform |
| Default | None |
| Description | Controls the initial values of the Q-tables. **None** fills the Q-tables with 0s. **Normal** fills the Q-tables with values from a normal distribution from -10 to 10. **Uniform** fills the Q-tables with values from a uniform distribution from -10 to 10. |

* **Opposition**

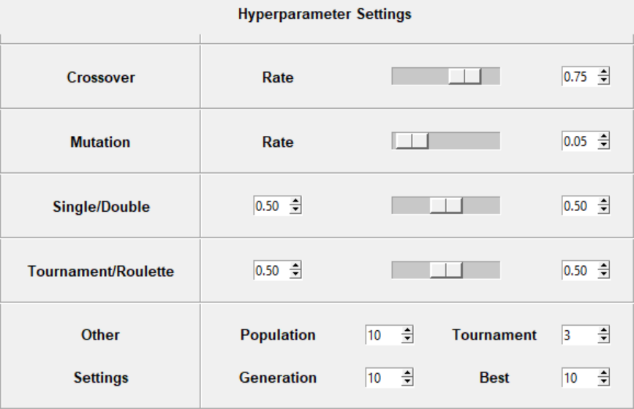
|  |  |
| --- | --- |
| Options | False, True |
| Default | False |
| Description | Allows the agent to perform opposition-based learning. The agent also takes the opposite action and updates the Q-table. The do-nothing action has no opposite action and is only updated once. |

* **Reward**

|  |  |
| --- | --- |
| Options | Constant, Time-Based,  X-Distance,  X-Distance (Center),  XY-Distance |
| Default | Constant |
| Description | Controls the reward function used by the agent. **Constant** returns 1 for each turn agent has not lost and 0 if the agent loses.  **Time-Based** returns more reward the more turns the agent has survived and 0 if the agent loses the game.  **X-Distance** returns a greater reward the shorter the horizontal distance between the midpoints of the ball and the paddle.  **X-Distance (Center)** returns a greater reward the shorter the horizontal distance between the midpoint of the ball and the entire paddle. A ball above the paddle will receive the maximum reward possible.  **XY-Distance** returns a greater reward the shorter the Euclidean distance between the midpoints of the ball and the paddle. |

## Hyperparameter Settings

The hyperparameter settings allow users to control the process of hyperparameter tuning of the agent. It is accessed by pressing the tuning button. The options available are shown below in Figure 3.



***Figure 3: Hyperparameter Settings***

* **Crossover Rate**

|  |  |
| --- | --- |
| Range | 0.00 – 1.00 |
| Default | 0.75 |
| Description | Controls the percentage of chromosomes that undergo the process of crossover. |

* **Mutation Rate**

|  |  |
| --- | --- |
| Range | 0.00 – 1.00 |
| Default | 0.05 |
| Description | Controls the percentage of genes that undergo the process of mutation. |

* **Single / Double**

|  |  |
| --- | --- |
| Range | 0.00 – 1.00 |
| Default | 0.50 / 0.50 |
| Description | Controls the percentage of single and double crossover when successful crossover occurs. |

* **Tournament / Roulette**

|  |  |
| --- | --- |
| Range | 0.00 – 1.00 |
| Default | 0.50 / 0.50 |
| Description | Controls the percentage of tournament and roulette selection for selection of the next generation. |

* **Population**

|  |  |
| --- | --- |
| Range | 10 – 100 |
| Default | 10 |
| Description | Controls the number of chromosomes in each generation. Can only be a multiple of 2. |

* **Elite**

|  |  |
| --- | --- |
| Range | 0 – 10 |
| Default | 2 |
| Description | Controls the number of fittest chromosomes that are guaranteed to survive to the next generation. Can only be a multiple of 2. |

* **Generation**

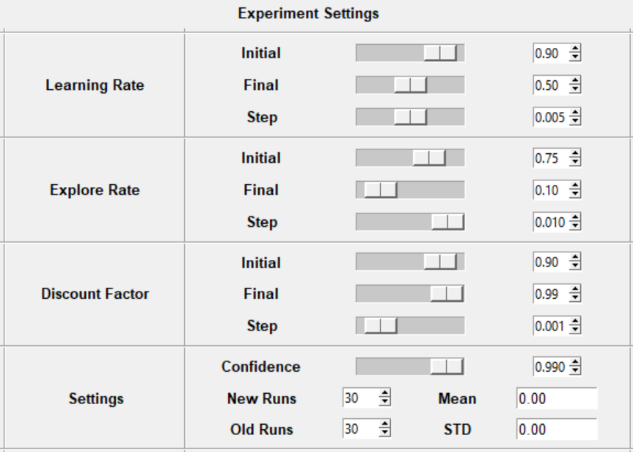
|  |  |
| --- | --- |
| Range | 1 – 100 |
| Default | 50 |
| Description | Controls the number of iterations of genetic algorithm to be run. |

* **Best**

|  |  |
| --- | --- |
| Range | 1 – 10 |
| Default | 10 |
| Description | Controls the number of fittest chromosomes saved and displayed at the end. |

## Experiment Settings

The experiment settings allow users to control the process of comparing different parameter settings. It is accessed by pressing the experiment button. The options available are shown below in Figure 4.



***Figure 4: Experiment Settings***

* **Learning Rate**

|  |  |
| --- | --- |
| Range | 0.00 – 1.00 / 0.000 – 0.010 |
| Default | 0.90 / 0.10 / 0.010 |
| Description | Controls how much the new learned value is used. |

* **Explore Rate**

|  |  |
| --- | --- |
| Range | 0.00 – 1.00 / 0.000 – 0.010 |
| Default | 0.50 / 0.01 / 0.010 |
| Description | Controls the chance that a random action is taken instead of the best action. |

* **Discount Factor**

|  |  |
| --- | --- |
| Range | 0.00 – 1.00 / 0.000 – 0.010 |
| Default | 0.90 / 0.99 / 0.001 |
| Description | Determines the importance of future rewards. |

* **Confidence**

|  |  |
| --- | --- |
| Range | 0.500 – 0.999 |
| Default | 0.990 |
| Description | Controls the confidence level of the t-test used to evaluate the results of the different experiments. |

* **New Runs**

|  |  |
| --- | --- |
| Range | 30 – 100 |
| Default | 30 |
| Description | Controls the number of successful runs needed for results collection. |

* **Old Runs**

|  |  |
| --- | --- |
| Range | 30 – 100 |
| Default | 30 |
| Description | Controls the size of the old sample to be compared. |

* **Mean**

|  |  |
| --- | --- |
| Range | 1.00 – 200.00 |
| Default | 45.63 |
| Description | Controls the mean of the old sample to be compared. |

* **STD**

|  |  |
| --- | --- |
| Range | 0.00 – 140.71 |
| Default | 1.40 |
| Description | Controls the standard deviation of the old sample to be compared. |

Experiment

The experiments to be run using the GUI are listed below. There are 3 main steps in the experiment listed below.

## Hyperparameter Tuning

The default settings are used for hyperparameter tuning. The genetic algorithm is run for 100 generations with a population size of 10. The best 10 chromosomes are displayed after each generation. The results of the final generation are compared before tuning the hyperparameters for the other experiments.

## Parameter Settings

The results of the default settings are used as the baseline for comparison. After that, each parameter is changed individually and the results recorded. Finally, the best settings of each parameter are combined to attempt to produce the best results.

## Game Settings

The results of the default settings are also used as the baseline for comparison. After that, each game setting is changed individually and the results recorded to study the effects of each game setting.

Results

The results of each experiment are documented below.

## Hyperparameter Tuning

There are lots of repetition in the rankings due to the different fitness scores for the same chromosome. So, only the top occurrence of genes is shown.

* **Hyperparameters**

|  |  |  |  |
| --- | --- | --- | --- |
| No | Learning | Explore | Discount |
| 1 | 82/92/1.0 | 19/11/0.7 | 36/37/0.7 |
| 2 | 82/92/1.0 | 06/13/0.6 | 87/95/0.7 |
| 3 | 82/92/1.0 | 06/11/0.7 | 36/37/0.7 |
| 4 | 50/48/0.1 | 29/11/0.3 | 36/25/0.0 |
| 5 | 82/77/1.0 | 29/11/0.7 | 36/15/0.7 |
| 6 | 50/48/1.0 | 29/11/0.7 | 36/15/0.7 |
| 7 | 82/48/1.0 | 29/11/0.7 | 04/15/0.7 |
| 8 | 82/48/1.0 | 32/11/0.3 | 36/97/0.7 |
| 9 | 82/48/1.0 | 06/11/0.7 | 36/37/0.7 |
| 10 | 82/48/1.0 | 29/11/0.3 | 36/37/0.7 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No | Mean | STD | Diff | Fail |
| 1 | 1.83 | 0.65 | 43.12 | 0 |
| 2 | 2.63 | 1.67 | 42.04 | 0 |
| 3 | 2.30 | 0.95 | 42.59 | 0 |
| 4 | 4.30 | 2.41 | 40.11 | 0 |
| 5 | 3.90 | 1.77 | 40.74 | 0 |
| 6 | 4.77 | 2.24 | 39.71 | 0 |
| 7 | 3.73 | 2.55 | 40.63 | 0 |
| 8 | 7.60 | 4.67 | 35.90 | 0 |
| 9 | 2.37 | 1.22 | 42.45 | 0 |
| 10 | 3.47 | 2.27 | 41.00 | 0 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No | Mean | STD | Diff | Fail |
| 1 | 1.83 | 0.65 | 43.12 | 0 |
| 3 | 2.30 | 0.95 | 42.59 | 0 |
| 9 | 2.37 | 1.22 | 42.45 | 0 |
| 2 | 2.63 | 1.67 | 42.04 | 0 |
| 10 | 3.47 | 2.27 | 41.00 | 0 |
| 5 | 3.90 | 1.77 | 40.74 | 0 |
| 7 | 3.73 | 2.55 | 40.63 | 0 |
| 4 | 4.30 | 2.41 | 40.11 | 0 |
| 6 | 4.77 | 2.24 | 39.71 | 0 |
| 8 | 7.60 | 4.67 | 35.90 | 0 |

|  |  |  |  |
| --- | --- | --- | --- |
| No | Learning | Explore | Discount |
| 1 | 82/92/1.0 | 19/11/0.7 | 36/37/0.7 |
| 3 | 82/92/1.0 | 06/11/0.7 | 36/37/0.7 |
| 9 | 82/48/1.0 | 06/11/0.7 | 36/37/0.7 |
| 2 | 82/92/1.0 | 06/13/0.6 | 87/95/0.7 |
| 10 | 82/48/1.0 | 29/11/0.3 | 36/37/0.7 |
| 5 | 82/77/1.0 | 29/11/0.7 | 36/15/0.7 |
| 7 | 82/48/1.0 | 29/11/0.7 | 04/15/0.7 |
| 4 | 50/48/0.1 | 29/11/0.3 | 36/25/0.0 |
| 6 | 50/48/1.0 | 29/11/0.7 | 36/15/0.7 |
| 8 | 82/48/1.0 | 32/11/0.3 | 36/97/0.7 |

## Parameter Settings

The results of the default settings are used as the baseline for comparison. Each parameter setting is first compared individually before combining.

* Q-Table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table | Mean | STD | Diff | Fail |
| 1 | 45.63 | 1.40 | - | 0 |
| 2 | 42.47 | 4.46 | 1.12 | 3 |
| 3 | 40.30 | 3.90 | 3.52 | 5 |
| 4 | 38.93 | 2.29 | 5.52 | 7 |
| 5 | 37.13 | 2.27 | 7.33 | 49 |

* State

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| State | Mean | STD | Diff | Fail |
| 1 | 45.63 | 1.40 | - | 0 |
| 2 | 43.07 | 3.26 | 1.01 | 0 |
| 3 | 40.17 | 3.61 | 3.77 | 2 |
| 4 | 38.13 | 3.09 | 6.01 | 0 |
| 5 | 37.60 | 2.88 | 6.63 | 0 |
| 6 | 35.10 | 3.99 | 8.68 | 0 |
| 7 | 33.53 | 3.12 | 10.60 | 0 |
| 8 | 30.83 | 4.19 | 12.87 | 0 |
| 9 | 32.17 | 3.59 | 11.78 | 0 |
| 10 | 31.00 | 3.89 | 12.83 | 0 |

* Action

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Action | Mean | STD | Diff | Fail |
| 2 | 45.63 | 1.40 | - | 0 |
| 3 | 44.23 | 1.30 | 0.56 | 0 |

* Random

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Random | Mean | STD | Diff | Fail |
| None | 45.63 | 1.40 | - | 0 |
| Normal | 45.57 | 1.57 | 0.00 | 0 |
| Uniform | 45.23 | 1.61 | 0.00 | 0 |

* Opposition

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Oppo | Mean | STD | Diff | Fail |
| 0 | 45.63 | 1.40 | - | 0 |
| 1 | - | - | - | - |

* Reward

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Func | Mean | STD | Diff | Fail |
| C | 45.63 | 1.40 | - | 0 |
| XC | 36.87 | 3.01 | 7.31 | 0 |
| T | 45.40 | 3.11 | 0.00 | 0 |
| X | 48.23 | 1.77 | 0.00 | 0 |
| XY | 76.87 | 4.20 | 0.00 | 3 |

* Combination

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Comb | Mean | STD | Diff | Fail |
| One | | | | |
| R | 36.87 | 3.01 | 7.31 | 0 |
| Two | | | | |
| RS | 29.40 | 2.62 | 14.93 | 0 |
| RO | 31.43 | 3.79 | 12.43 | 0 |
| RA | 39.97 | 3.47 | 4.03 | 0 |
| RQ | 40.10 | 3.53 | 3.87 | 0 |
| Three | | | | |
| RSO | 29.63 | 4.94 | 13.75 | 0 |

## Game Settings

The results of the default settings are used as the baseline for comparison. Each game setting is compared individually.

* Brick Placement

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Brick | Mean | STD | Diff | Fail |
| Row | 45.63 | 1.40 | - | 0 |
| Col | 45.77 | 0.86 | 0.00 | 1 |
| Rand | 46.33 | 0.80 | 0.00 | 0 |

* Rows

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Row | Mean | STD | Diff | Fail |
| 1 | 41.40 | 1.30 | 3.39 | 0 |
| 2 | 44.97 | 1.27 | 0.00 | 0 |
| 3 | 45.63 | 1.40 | - | 0 |
| 4 | 46.77 | 0.77 | 0.00 | 0 |
| 5 | 47.43 | 0.77 | 0.00 | 1 |
| 6 | 47.70 | 0.47 | 0.00 | 1 |

* Columns

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Col | Mean | STD | Diff | Fail |
| 1 | 33.53 | 4.60 | 10.00 | 0 |
| 2 | 37.27 | 1.51 | 7.46 | 0 |
| 3 | 38.73 | 1.51 | 6.00 | 0 |
| 4 | 42.40 | 1.73 | 2.26 | 0 |
| 5 | 43.57 | 1.76 | 1.08 | 0 |
| 6 | 44.47 | 1.22 | 0.35 | 0 |
| 8 | 45.63 | 1.40 | - | 0 |
| 10 | 46.60 | 0.67 | 0.00 | 0 |
| 12 | 47.13 | 0.73 | 0.00 | 1 |
| 15 | 47.70 | 0.65 | 0.00 | 1 |

* Ball Speed

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Speed | Mean | STD | Diff | Fail |
| 1 | 49.83 | 1.02 | 0.00 | 0 |
| 2 | 48.80 | 0.41 | 0.00 | 0 |
| 3 | 47.87 | 0.51 | 0.00 | 0 |
| 4 | 46.67 | 0.71 | 0.00 | 0 |
| 5 | 45.63 | 1.40 | - | 0 |
| 6 | 45.43 | 0.94 | 0.00 | 0 |
| 7 | 44.63 | 1.63 | 0.06 | 0 |
| 8 | 44.57 | 1.38 | 0.20 | 5 |
| 9 | 46.53 | 1.28 | 0.00 | 21 |
| 10 | 48.27 | 1.28 | 0.00 | 89 |

* Paddle Speed

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Speed | Mean | STD | Diff | Fail |
| 5 | 47.37 | 1.27 | 0.00 | 22 |
| 6 | 45.67 | 0.96 | 0.00 | 0 |
| 7 | 45.87 | 1.14 | 0.00 | 0 |
| 8 | 45.30 | 1.66 | 0.00 | 0 |
| 9 | 45.70 | 1.24 | 0.00 | 0 |
| 10 | 45.63 | 1.40 | - | 0 |
| 11 | 45.83 | 1.15 | 0.00 | 0 |
| 12 | 45.70 | 0.99 | 0.00 | 0 |
| 13 | 45.93 | 1.36 | 0.00 | 0 |
| 14 | 46.27 | 1.08 | 0.00 | 0 |
| 15 | 45.90 | 0.84 | 0.00 | 1 |

* Game Mode

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Mode | Mean | STD | Diff | Fail |
| 0 | 45.63 | 1.40 | - | 0 |
| 1 | 45.63 | 1.40 | 0.00 | 0 |

Discussion

The discussion of the results each experiment is shown below.

## Hyperparameter Tuning

This part discusses about

## Parameter Settings

This part

* Q-Table
* State
* Action
* Random
* Opposition
* Reward
* Combination

## Game Settings

This part

* Brick Placement
* Rows
* Columns
* Ball Speed
* Paddle Speed
* Game Mode