Dueling DQN

Liran Nochumsohn 204693410 and Yair Vaknin 034662072.

In this project implemented the Dueling DQN function approximation as proposed by DeepMind research group in the Dueling Network Architectures for Deep Reinforcement Learning Paper. The method presents state-of-the art performance improvement on many Atari 2600 game benchmarks that were used for the original Human-Control paper.

To attempt to achieve similar results we used the same methods proposed by the research paper starting from preprocessing neural network architecture and hyperparameter tuning. In this we will present our methods and results.

Preprocessing

For this project we used the gym environments for testing, specifically the game Breakout. Gym offers at state space represented with 210x160x3 images. Our goal is to make the results as similar as possible to DeepMind's results and faster computation, hence we transformed the images and resized them to 84x84x3 images where 3 represents the pixel color as RGB.

Dueling DQN Implementation and Parameters

Most attributes of the Dueling DQN are taken from the Human-Control paper by DeepMind therefore importing the parameters and architecture from the proposed DQN in Human-control was crucial to the implementation. We gave a lot of importance to the Human – Control paper because it operated as the baseline framework.

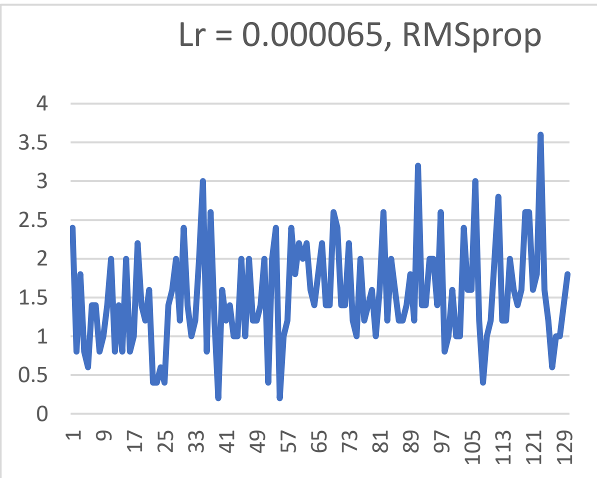
In our testing we tested different combinations of hyperparameters focusing mostly on the learning rate of the optimization function as well as the type of the optimization function. Focusing on a larger range of hyperparameters was simply not feasible due to the computational limitations of our Hardware. The following parameters were considered in different combinations:

- Learning Rates : 0.001, 0.0005, 0.00025, 0.0001, 0.0005, 0.00001
- Optimization Functions: RMSprop, Stochastic Gradient Descent and Adam.

The rest of the parameters we chosen based on Human-Control.

Results

Even after days of learning with different variation we could not reach sufficient results that demonstrate an obvious improvement. Our best results came from the run the included the exact hyperparameters offered by The Dueling DQN paper.



2-episode average performance.

The X axis represents every second episode and y axis represents the total rewards

One can see a slight improvement, but no significance. This be as a result of wrong hyperparameters or not sufficient learning time.

The amount of learning time was limited for each learning attempt to find the ideal hyperparameter combination in a limited time span and limited computational power.

Implementation

Learning_new file is comprised of the functions used for learning and evaluating the performance of the algorithm. Running the program will results in learning new parameters for the neural network for solving the game. The program is designed to run a checkpoint even two episodes. The checkpoint save the neural networks parameters and evaluation values relevant for understanding the progress such as computation time, average rewards and reward standard deviation.

All runs and evaluations were tested on the game Breakout. The reason the game was chosen was because Breakout was one of the game chosen for hyperparameter tuning in the Human-Control paper.