AI Agent for NES Super Mario Brothers

Denis Barkar (dbarkar@stanford.edu)

Motivation and Goals

Super Mario Bros is a 2D platformer game released for NES in 1985. The objective is to complete first level by traversing the terrain rightward without dying.

Problem definition and challenges

OpenAI Gym Retro toolkit is used.
Actions: Left, Right, Up, Down, A, B.
Observation: RGB image, an array of shape $[240, 224, 3]$.
Reward: returned by the environment, clipped to be in the range $[-1; 1]$.
Evaluation metric is the total reward that the agent collects in an episode (the final score).
Main challenge is that the score doesn't reflect real progress in the game.

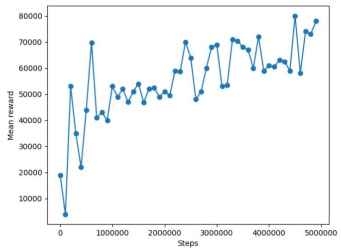
Approach

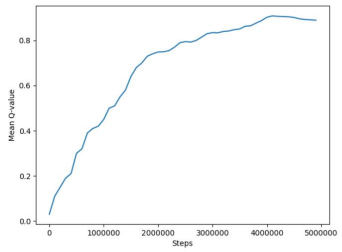
I used Double Deep Q-learning [1] with prioritized experience replay where we optimize the loss function with stochastic gradient descent.
The image frame is cropped, downscaled, converted to grayscale and normalized to an 84×84 black and white image. Each state consists of 4 frames.
The reward r_t is determined by a linear combination of the change in the total game score, the distance the agent moved to the right, number of lifes, number of coins and current Mario state (small, big, with fireball).
The convolutional neural network architechture:

1. Input: Four grayscale frames with resolution of 84×84 pixels.
2. Hidden layer: Convolves 8 7×7 filters of stride 2 with the input image and applies a rectifier nonlinearity
3. Hidden layer: Max pooling 3×3 of stride 2
4. Hidden layer: Convolves 16 5×5 filters of stride 2 and applies a rectifier nonlinearity
5. Hidden layer: Max pooling 3×3 of stride 2
6. Hidden layer: Convolves 32 3×3 filters of stride 2 and applies a rectifier nonlinearity
7. Hidden layer: Fully connected layer that consists of 256 rectifier unit
8. Output: Fully connected linear layer which outputs Q-values of each valid action (6 actions in total)

After some experiments [2], initial adaptive learning rate α was set to 0.0005, discount rate γ to 0.95. An epoch was set at 200.000 learning steps.

Results





An evaluation routine is called every epoch to evaluate the progress of the agent using parameters

- the mean Q-value, defined as the average maximum value of Q for all states in the network
- the mean score in the game achieved during the evaluation process

Many hours were spent to fine tune hyper-parameters and reward function.
The training was done for a total of 5000000 steps using NVIDIA RTX 2080 Ti video card. Results can be seen in figures.
Agent's performance was good enough to finish first level of the game.
Although the following approach worked, its results are not very encouraging, probably due to NES games superior complexity against Atari 2600 games.
The recording can be found here: `https://youtu.be/BwjKn_A--vk`.

Analysis and Future Work

In this project, I designed an AI agent using Double Deep Q-learning in order to play NES Super Mario Bros which was able to successfully complete first level of the game.
For future work, it is possible to:

- increase the number of the network's layers
- train the network for much more time
- try Actor-Critic algorithm and its modifications
- try exploration based algorithm.

References

[1] Hado van Hasselt, Arthur Guez, and David Silver. Deep reinforcement learning with double q-learning. *CoRR*, abs/1509.06461, 2015.

[2] Vincent François-Lavet, Raphaël Fonteneau, and Damien Ernst. How to discount deep reinforcement learning: Towards new dynamic strategies. *CoRR*, abs/1512.02011, 2015.

[3] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin A. Riedmiller. Playing atari with deep reinforcement learning. *CoRR*, abs/1312.5602, 2013.

[4] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, February 2015.