ECE239AS Project Proposal

Dueling Network Architectures for Deep Reinforcement Learning

Sarat Bhargava
805219546
sarat2895@ucla.edu

Eden Haney
005221845
ehaney@ucla.edu

Radhika Nayar
805226085
radhica@ucla.edu

Abstract

Most approaches for improving Deep RL's performance focus on the learning algorithm itself while using standard Neural Networks not necessarily designed for reinforcement learning. Here we propose to study and implement a potential improvement to the standard neural networks used in Deep Q-learning applications, known as the dueling network architecture. We expect this network to outperform classic implementations of other Deep Q-learning algorithms in both policy evaluation and Atari game environments.

1 Background

Deep Reinforcement Learning (Deep RL) is an exciting topic due to many recent successes. However, most of the approaches for Deep RL use standard neural networks, such as convolutional networks, multilayer perceptrons (MLPs), long short-term memory networks (LSTMs) and autoencoders. The main focus in recent advances has been on designing improved RL algorithms, or simply on incorporating existing neural network architectures into RL methods. Instead of improving the learning algorithm itself, we propose to address how performance can be improved by building a neural network architecture that is customized for model-free RL.

2 Project Focus

For this project, we will understand, survey and implement a neural network architecture for model-free reinforcement learning known as the *dueling architecture* (1). This architecture, visualized in figure 1, advances a new network by improving already published algorithms such as the already robust Double Deep Q-learning (Double DQN) algorithm (2) and Deep Q-Networks (DQN) (3). Thus, a true understanding of the dueling architecture requires a solid understanding of the algorithms on which it is built.

The dueling architecture is a single Q-network architecture where the lower layers are convolutional, as in DQN. However, instead of following the convolutional layers with a single sequence of fully connected layers, we use two sequences (or streams) of fully connected layers. The constructed streams provide separate estimates: one for estimating the state-value function and other for estimating state-dependent action advantage function. The two streams are then combined to produce a single output Q-function. Similar to DQN, the dueling architecture outputs a set of Q-values, one for each action. We will be tweaking the OpenAI baselines implementation of DQN, dueling DQN to our task.

UCLA ECE 239AS Reinforcement Learning Spring 2020.

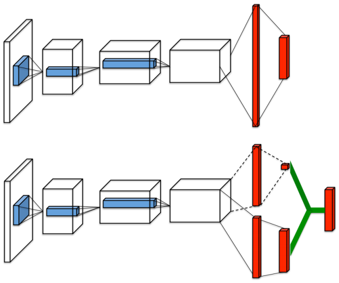

Figure 1: A popular single stream Q-network (top) and the dueling Q-network (bottom). The dueling network has two streams to separately estimate state-values and the advantages for each action; the green output module combines them. Both networks output Q-values for each action. From (1).

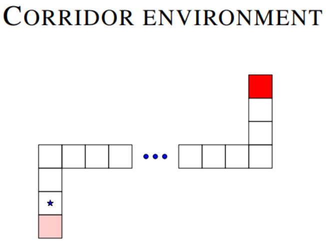

Figure 2: The corridor environment. The redness of a state signifies the reward the agent receives upon arrival. The game terminates upon reaching either reward state. The agent's actions are going up, down, left, right and no action. From (1).

A key motivator behind this architecture is that for some games, it is unnecessary to know the value of each action at every time step. By explicitly separating two estimators, the dueling architecture can learn which states are valuable, without having to learn the effect of each action for each state.

3 Evaluation

We propose two separate and complimentary simulations to test the architecture on. We will first examine a corridor simulation, which allows a focus on policy evaluation, and then use the Space Invaders Atari environment from OpenAI Gym.

The corridor environment (figure 2) is composed of three connected corridors, where each of the two vertical sections have 10 states while the horizontal section has 50. The agent will start from the bottom left corner of the environment and must move to the top right to get the largest reward. A total of 5 actions are available: go up, down, left, right, and no-op. There is freedom of adding an arbitrary number of no-ops actions.

Next, we will evaluate our architecture on an Arcade Learning Environment which is composed of Atari games. The challenge will be to learn to play one of the Atari games (such as Space Invaders) given only raw pixel observations and game rewards.

In this project, we want to demonstrate the superior performance of the dueling architecture compared to the previous published algorithms such as DQN or Double DQN. To evaluate the simple policy evaluation task in the corridor environment, we will evaluate a single-stream Q-architecture (DQN) against the dueling architecture on three variants of the corridor environment with 5, 10, and 20 actions respectively. We will measure the performance by squared error against the true state values using equation 1.

$$\sum_{s\epsilon S, a\epsilon A} (Q(s,a;\theta) - Q^{\pi}(s,a))^2 \tag{1}$$

We will then plot the squared error for policy evaluation with 5,10, and 20 actions on log-log scale for both the single-stream network (DQN) and the dueling network. We expect to find the dueling network to consistently outperform a conventional single-stream network, with the performance gap increasing with the number of actions.

To evaluate the performance in the general Atari game environment, we will measure improvement by the percentage change in score over the better of human and baseline agent scores using the metric described in equation 2.

$$\frac{Score_{Agent} - Score_{Baseline}}{max(Score_{Human}, Score_{Baseline}) - Score_{Random}} \tag{2}$$

We will plot the results, but expect the dueling network to considerably outperform the single-stream network.

References

[1] Z. Wang, T. Schaul, M. Hessel, H. van Hasselt, M. Lanctot, and N. de Freitas, “Dueling network architectures for deep reinforcement learning,” *arXiv:1511.06581 [cs]*, Apr 2016. arXiv: 1511.06581.

[2] H. van Hasselt, A. Guez, and D. Silver, “Deep reinforcement learning with double q-learning,” *arXiv:1509.06461 [cs]*, Dec 2015. arXiv: 1509.06461.

[3] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, and et al., “Human-level control through deep reinforcement learning,” *Nature*, vol. 518, p. 529–533, Feb 2015.