Continuous versus discrete action spaces for deep reinforcement learning

Julius Stopforth[1,2] and Deshendran Moodley[1,2]

[1] University of Cape Town
[2] Center for Artificial Intelligence Research

Abstract. Reinforcement learning problems may have either a discrete or continuous action space that greatly affects the algorithm used. Deep reinforcement learning algorithms have already been applied to both discrete and continuous action spaces. In this work we compare the performance of two well established model-free DRL algorithms: Deep Q Network for discrete action spaces, and the continuous action space variant Deep Deterministic Policy Gradient on the same RL problem of the LunarLander. Furthermore, we investigate to what extent Experience Replay affects the comparative performance of both algorithms for limited training times.

Keywords: reinforcement learning, continuous control, deep neural networks

1 Introduction

In this work, we attempt to compare the effect of discrete and continuous action spaces on the training of a deep reinforcement learning agent. Specifically, we look at the performance of the well established Deep Q-Network (DQN) algorithm[3] compared to its continuous action space variant the Deep Deterministic Policy Gradient (DDPG) algorithm[2].

The research aims to determine if and or when there are distinct advantages to using discrete or continuous action spaces when designing new DRL problems and algorithms. In this work we present preliminary results for both the DQN and DDPG algorithms to a known RL problem of the LunarLander using OpenAI Gym[1]. By comparing the performance of the aforementioned algorithms in a known environment, we hope to gain insight into how the difference between continuous and discrete action spaces affects the training and performance of these algorithms.

2 Experiments

The LunarLander environment provided by OpenAI already has two variants for both discrete and continuous action spaces and was used without modification.

The LunarLander is considered "solved" when the algorithm achieves an average reward of 200 points on 100 independent trials.

Each algorithm is given 100, 200, and 500 episodes to train before measuring the average reward over 100 independent trials. The experiements were repeated 10 times each in order to eliminate the possibility of a singularly excellent result and facilitates the aim of comparative analysis between the two variations in the algorithms used.

Both algorithms were implemented the same network structure of a single fully connected hidden layer of 10 nodes. The network structures used ReLU activation layers and the RMSProp optimiser. Huber loss was used for the algorithms. The learning rate and greediness of both algorithms was also kept the same.

3 Results

In comparison to the DQN algorithm, the DDPG algorithm performed worse over 500 training episodes.

Table 1. Average reward for 100 trials for the DQN with experience replay

No. training episodes	**100**	**200**	**500**
Average reward	-430	-420	-420

Table 2. Average reward for 100 trials for the DDPG with experience replay

No. training episodes	**100**	**200**	**500**
Average reward	-733	-687	-718

4 Discussion

The preliminary results presented in this work are align with the results obtained from the HEDGER algorithm[4] and suggest that DQN outperforms DDPG when the number of training episodes is limited. However, the results presented are inconclusive and limited.

Ongoing work includes extending the number of training episodes as well as the increasing the complexity of the deep learning structures used in order to gain deeper insight into the performance of the algorithms.

References

1. Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J., Zaremba, W.: Openai gym (2016)
2. Lillicrap, T.P., Hunt, J.J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., Wierstra, D.: Continuous control with deep reinforcement learning. arXiv preprint arXiv:1509.02971 (2015)
3. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., et al.: Human-level control through deep reinforcement learning. Nature **518**(7540), 529 (2015)
4. Smart, W.D., Kaelbling, L.P.: Practical reinforcement learning in continuous spaces. In: ICML. pp. 903–910 (2000)