**Continuous control in deep reinforcement learning with direct policy derivation from Q network**

Aydar Akhmetzyanov, Alexandr Klimchik, Rauf Yagfarov, Salimzhan Gafurov, Mikhail Ostanin

Innopolis University, Innopolis, 420500, Russia

Reinforcement learning approach allows to learn desired control policy in different environments without explicitly providing system dynamics. Model-free deep Q-learning algorithm is proven to be efficient on a large set of discrete-action tasks. Extension of this method to the continuous control task usually solved with actor-critic methods which approximate a policy function with additional actor network and uses Q function to speed up policy network training. Another approach is to discretize action space which will not give a smooth policy and is not applicable for large action spaces. A direct continuous policy derivation from Q network leads to optimization of action on each inference and training step which is not efficient but provides the optimal and continuous action. Time-efficient Q function input optimization is required in order to apply this method in practice. In this work we implement efficient action derivation method which allows using Q learning in real time continuous control tasks. In addition, we test our algorithm on robotics control tasks from robotics gym environments and compare this method with modern continuous RL methods. The results have shown that in some cases proposed approach learns smooth continuous policy keeping the implementation simplicity of the original discreet action space Q learning algorithm.

Reinforcement learning Q-learning Continuous control Robotics Deep learning

1. Introduction

1.2 State of the art algorithms overview

1. Q learning

2.1 Q learning optimizations

1. Continuous control with direct policy derivation from Q network
   1. method description
   2. network architecture(same as for baselines)
   3. Hyperparameter description
   4. policy performance
   5. sample efficiency
   6. behavioral analysis and policy efficiency
   7. further improvements
2. Conclusion

Acknowledgments. This research was supported by the Russian Education Ministry under the Agreement No. 075-10-2018-017 on granting a subsidy from 11.26.2018 “Development of a commercial urban transport with the intellectual driver assistance system “City Pilot”. The customer is a federal target program. The unique identifier of the project RFMEFI60918X0005. Industrial partner - KAMAZ

References