UNIVERSITY
OF TECHNOLOGY
SYDNEY

Sample-efficient deep reinforcement learning from single agent to multiple agents

by Han Zheng

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

under the supervision of Dr. Jing Jiang, A/Prof. Guodong Long, and Prof. Chengqi Zhang

University of Technology Sydney
Faculty of Engineering and Information Technology

September 2021

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, Han Zheng declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution. This research is supported by the Australian Government Research Training Program.

September,2021 Han Zheng

Production Note:
Signature removed
prior to publication.

ABSTRACT

Deep reinforcement learning (DRL) has recently become a very popular topic in the academic field. However, it usually suffers the sample inefficiency problem due to the lack of effective exploration, instability, or temporal credit assignment issue. High sample complexity leads to a huge computation cost and adversely affects the employment of DRL techniques in practice. Despite many methods proposed to address this challenge, further improvements are still needed. This thesis contributes to developing sample-efficient DRL methods for continuous control from two perspectives: single agent and multiple agents. Specifically, the key contribution includes an uncertainty regularized policy learning method for single agent and two ensemble learning frameworks for multiple agents. Importantly, this thesis highlights that the multiple agents' methods can be seen as bridging gaps among on-policy, off-policy RL, and evolutionary algorithms. Moreover, our approach achieves consistent improvements over the baseline methods and gives novel insight into effectively taking advantage of different methods to get the best of them.

Dissertation directed by Dr. Jing Jiang, A/Prof. Guodong Long, Prof. Chengqi Zhang

Australian Artificial Intelligence Institute

Faculty of Engineering and IT

University of Technology Sydney

Acknowledgements

Firstly, I would like to thank my academic supervisor Dr. Jing Jiang, A/Prof. Guodong Long, and Distinguished Professor Chengqi Zhang. I am incredibly grateful for their kind, timely, and strong support. Dr. Jing Jiang is very nice and patient in guiding me on the research trip. When I was stuck on my first paper for a long time, Dr. Jing Jiang gave me the in-time support to help me out of the woods. Under her guidance, I gradually understand how to do research properly, which is of significant importance for my later works. I appreciate it very much. A/Prof. Guodong Long is the one who brings me into the academic career. When I was working in a company and did something that I did not like very much, he gave me an opportunity to do something different. Thanks for his belief. I am also very thankful for them for giving me the freedom to choose interesting research topics. Prof. Chengqi Zhang provides me with a resourceful platform and teaches me how to do research and live a meaningful life from a bigger picture. Thanks for his inspiring words.

I would also want to thank Dr. Pengfei Wei for providing me guidance and detailed suggestions on my research. I still remember the time when we have an intense discussion on paper writing. Thanks for his patience, motivation, and immense knowledge. My sincere thanks also go to Prof. Xuan Song for allowing me to join their team as an intern.

I would also like to thank my lab-mates at UTS. Without their help, I cannot get used to the new environment soon, both in my research and daily life. Last but not the least, I would like to thank my parents. They are always my strong backing. Love you forever.

Han Zheng

Sydney, Australia, 2021.

List of Works

C-1. **Han Zheng**, Jing Jiang,Pengfei Wei,Guodong Long and Chengqi Zhang, "Competitive and Cooperative HeterogReinforcement Learning". *Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems(AAMAS 2020)* (CORE Rank A* Conference, 8 pages)

C-2. **Han Zheng**, Pengfei Wei, Jing Jiang, Guodong Long, Qinghua Lu and Chengqi Zhang. "Cooperative Heterogeneous Deep Reinforcement Learning". *34th Conference on Neural Information Processing Systems (NeurIPS 2020)* (CORE Rank A* Conference, 9 pages)

C-3. **Han Zheng**, Jing Jiang, Pengfei Wei, Guodong Long, Xuan Song and Chengqi Zhang. "Uncertainty Regularized Policy Learning for Offline Reinforcement Learning". *Draft to be submitted to ICLR 2022*

C-4. **Han Zheng**, Xufang Luo, Pengfei Wei, Xuan Song, Dongsheng Li and Jing Jiang. "Adaptive conservative Q-ensemble learning for offline-online interleaving learning". *Draft to be submitted to ICLR 2022*

Contents

List of Figures