

Self-Learning Agents in Stochastic Games

Mahbub khan

Declaration

Contents

1	Introduction	1
2	Problem Description	2
3	Literature & Tech Review	3
3.1	Literature Review	3
3.2	Tech Review	3
3.2.1	Reinforcement Learning and Markov Decision Process . .	3
3.2.2	Q-Learning	3
3.2.3	Policy Gradient	3
3.2.4	Actor-Critic Algorithms	3
4	Methods	4
4.0.1	Maximum Entropy RL	4
4.0.2	Soft Q-Learning and Soft Actor-Critic	4
5	Implementation	5
5.0.1	Pseudo Code: Soft Actor-Critic Algorithm	5
6	Results	6
6.0.1	Comparison between Deep Deterministic Policy Gradient(DDPG),Advantage Actor Critic(A2C), Soft Actor Critic(SAC)	6
7	Summary	7
8	Overview & Outlook	8

List of Figures

List of Tables

Abstract

Chapter 1

Introduction

Chapter 2

Problem Description

This is a reference [5],

Chapter 3

Literature & Tech Review

3.1 Literature Review

Main Papers to review [5],[4],[6],[3],[8],[1],[2],[7]

3.2 Tech Review

3.2.1 Reinforcement Learning and Markov Decision Process

3.2.2 Q-Learning

3.2.3 Policy Gradient

3.2.4 Actor-Critic Algorithms

Deep Deterministic Policy Gradient

Advantage Actor Critic

Chapter 4

Methods

4.0.1 Maximum Entropy RL

This is the paper I am using as theory [3]

4.0.2 Soft Q-Learning and Soft Actor-Critic

[1]
This is the algorithm I want to use [2]

Chapter 5

Implementation

5.0.1 Pseudo Code: Soft Actor-Critic Algorithm

Chapter 6

Results

6.0.1 Comparison between Deep Deterministic Policy Gradient(DDPG), Advantage Actor Critic(A2C), Soft Actor Critic(SAC)

Chapter 7

Summary

Chapter 8

Overview & Outlook

Bibliography

- [1] T. Haarnoja, H. Tang, P. Abbeel, and S. Levine. Reinforcement learning with deep energy-based policies. In D. Precup and Y. W. Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 1352–1361. PMLR, 06–11 Aug 2017.
- [2] T. Haarnoja, A. Zhou, K. Hartikainen, G. Tucker, S. Ha, J. Tan, V. Kumar, H. Zhu, A. Gupta, P. Abbeel, and S. Levine. Soft actor-critic algorithms and applications. *CoRR*, abs/1812.05905, 2018.
- [3] S. Levine. Reinforcement learning and control as probabilistic inference: Tutorial and review. *CoRR*, abs/1805.00909, 2018.
- [4] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. A. Riedmiller. Playing atari with deep reinforcement learning. *CoRR*, abs/1312.5602, 2013.
- [5] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, Feb. 2015.
- [6] B. O’Donoghue, R. Munos, K. Kavukcuoglu, and V. Mnih. PGQ: combining policy gradient and q-learning. *CoRR*, abs/1611.01626, 2016.
- [7] J. Oh, Y. Guo, S. Singh, and H. Lee. Self-imitation learning. *CoRR*, abs/1806.05635, 2018.
- [8] D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller. Deterministic policy gradient algorithms. In *Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32*, ICML’14, page I–387–I–395. JMLR.org, 2014.