INSTITUT FÜR INFORMATIK
Machine Learning

Universitätsstr. 1 D–40225 Düsseldorf

Actor-Critic Reinforcement Learning With Experience Replay

Julian Robert Ullrich

Bachelorarbeit

Beginn der Arbeit:	25. Juli 2018
Abgabe der Arbeit:	25. Oktober 2018
Gutachter:	Univ.-Prof. Dr. S. Harmeling
	Univ.-Prof. Dr. M. Leuschel

Erklärung

Hiermit versichere ich, dass ich diese Bachelorarbeit selbstständig verfasst habe. Ich habe dazu keine anderen als die angegebenen Quellen und Hilfsmittel verwendet.

Düsseldorf, den 25. Oktober 2018

Julian Robert Ullrich

Abstract

Deep Reinforcement Learning and policy gradient methods majorly contributed to the most recent advances in the field of Artificial Intelligence. These Methods enabled machines to surpass human performance for Atari console games (Mnih et al., 2015), boardgames like Chess, Shogi (Silver et al., 2017a) or Go (Silver et al., 2017b) and most recently even complex team-based computer games (OpenAI, 2018).

As environments grow in complexity, their simulation requires more computational ressources. Sample efficiency has therefore become an important aspekt of reinforcement learning.

The goal of this thesis is the implementation and evaluation of the "Actor-Critic with Experience Replay"(ACER) algorithm proposed by Wang et al., 2016 on the Atari console games.

Contents

1 Abstract **1**

2 Introduction **1**

2.1 Motivation/Objectives 1

3 Reinforcement Learning Frameworks **2**

3.1 Elements of Reinforcement Learning 2

3.2 Markov Decision Process 3

4 Actor-Critic Methods **4**

4.1 Actor-Only Methods 4

4.2 Critic-Only Methods 4

4.3 Actor-Critic Methods 5

4.4 A3C : Asynchronous Advantage Actor-Critic 5

5 Off-Policy Learning **6**

5.1 Importance-Sampling 6

5.2 Q-Retrace 6

6 ACER : Actor-Critic with Experience Replay **7**

7 Experimental Setup **8**

8 Experiments **9**

8.1 Hyperparameter Robustness 9

8.2 Adversial Attacks 9

8.3 Results 10

9 Conclusion **10**

10 References **11**

References **12**

List of Figures **13**

List of Tables **13**

1 Abstract

2 Introduction

2.1 Motivation/Objectives

3 Reinforcement Learning Frameworks

3.1 Elements of Reinforcement Learning

3.2 Markov Decision Process

4 Actor-Critic Methods

4.1 Actor-Only Methods

4.2 Critic-Only Methods

4.3 Actor-Critic Methods

4.4 A3C : Asynchronous Advantage Actor-Critic

5 Off-Policy Learning

5.1 Importance-Sampling

5.2 Q-Retrace

6 ACER : Actor-Critic with Experience Replay

7 Experimental Setup

8 Experiments

8.1 Hyperparameter Robustness

8.2 Adversial Attacks

8.3 Results

9 Conclusion

10 References

References

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis (2015). "Human-level control through deep reinforcement learning". In: *Nature* 518.7540, pp. 529–533.

OpenAI (2018). *OpenAI Five*.

David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharshan Kumaran, Thore Graepel, Timothy P. Lillicrap, Karen Simonyan, and Demis Hassabis (2017a). "Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm". In: *CoRR* abs/1712.01815. arXiv: 1712.01815.

David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, Yutian Chen, Timothy Lillicrap, Fan Hui, Laurent Sifre, George van den Driessche, Thore Graepel, and Demis Hassabis (2017b). "Mastering the game of Go without human knowledge". In: *Nature* 550, pp. 354–.

Ziyu Wang, Victor Bapst, Nicolas Heess, Volodymyr Mnih, Rémi Munos, Koray Kavukcuoglu, and Nando de Freitas (2016). "Sample Efficient Actor-Critic with Experience Replay". In: *CoRR* abs/1611.01224. arXiv: 1611.01224.

List of Figures

List of Tables