

Research Internship
École polytechnique
MSc&T Artificial Intelligence and Advanced Visual Computing
Academic Year 2018-2019

Safe Exploration with Simulator in Reinforcement Learning Algorithms

CONFIDENTIAL REPORT

Submitted by

Léon Zheng
Promotion X2016
École polytechnique

Under the guidance of

Iwane Hidenao
Internship tutor
AI Lab, Fujitsu Laboratories Ltd.

Erwan Scornet
Referent teacher
CMAP, École polytechnique



Artificial Intelligence Laboratory
FUJITSU LABORATORIES LTD.
1-1, Kamikodanaka 4-chome, Makahara-ku, Kawasaki
211-8588, Japan
Summer Internship (April 2019 - August 2019)

Declaration of Academic Integrity

Hereby I, Léon Zheng, confirm that:

1. the results presented in this report are my own work;
2. I am the author of this report;
3. I have not used the work of others without clearly acknowledging it, and quotations and paraphrases from any source are clearly indicated.

Léon Zheng
Promotion X2016

August 30th, 2019

Abstract

Deploying reinforcement learning algorithms in real life applications requires to take into account safety during the exploration process: some critical constraints must never be violated. In complementary to recent works about optimal policy search under explicit constraints in high-dimensional continuous state and action spaces, we introduce a first framework in which reinforcement learning algorithms can exploit external knowledge from a given simulator in order to do safe exploration. By controlling the behavior differences between dynamics in the simulator and in the real world, we build an algorithm which can evaluate the risk of an action by simulating it, so that the agent can know if this action is safe to execute it in the real world. In experiences where we consider tasks with linear and non-linear dynamics, like Linear Quadratic Regulator or Swimmer task, we illustrate the validity of this algorithm, in the sense that the agent never breaks given safety constraints during the whole training, when it uses correctly external knowledge from the simulator.

Contents

1	Introduction	1
1.1	RL algorithms on complex problems	1
1.2	RL algorithms in real life setting	2
1.3	Safe exploration with simulator in RL	2
2	Related work	4
2.1	Definitions and approaches of safety	4
2.2	Safe exploration with explicit constraints	5
3	Problem setting	7
3.1	Preliminaries in RL	7
3.1.1	Constrained Markov Decision Process	7
3.1.2	Policy and return	8
3.2	Definition of safety	9
3.3	Problem definition	10
4	Safe exploration with simulator	11
4.1	Model of the real world	11
4.1.1	Parameterized models	11
4.1.2	Real world parameter estimation	12
4.2	Safe exploration algorithm	13
4.2.1	Control of the simulator-reality gap	13
4.2.2	Algorithm	15
4.2.3	Analysis of the algorithm	16
5	Experiments	18
5.1	Reinforcement learning tasks	18
5.1.1	Linear dynamics	18
5.1.2	Non-linear dynamics	19
5.2	Baseline RL algorithms	21
5.2.1	Continuous Actor-Critic Learning Automaton	21

5.2.2	Augmented Random Search	21
5.3	Experiments results	22
6	Discussions	25
6.1	Conclusion	25
6.2	Future works	25
	Acknowledgements	27
	References	28
	Appendices	30
A	Swimmer task description	31
A.1	State and action variables	31
A.2	System dynamics	32
A.3	Typical values	32
B	Baseline RL algorithms	33
B.1	CACLA	33
B.2	ARS	34