THE UNIVERSITY OF EDINBURGH

INFORMATICS HONORS PROJECT

Generative Adversarial Networks Generating Novel Reinforcement Learning Policies

Author:
Giovanni ALCANTARA

Supervisor:
Dr. Timothy HOSPEDALES

A thesis submitted in fulfillment of the requirements for the degree of Bachelor's of Engineering

in the

School of Informatics
College of Science & Engineering

January 24, 2018

THE UNIVERSITY OF EDINBURGH

Abstract

College of Science & Engineering

Bachelor's of Engineering

Generative Adversarial Networks Generating Novel Reinforcement Learning Policies

by Giovanni ALCANTARA

The Thesis Abstract is written here (and usually kept to just this page). The page is kept centered vertically so can expand into the blank space above the title too...

Contents

Abstract iii

1 Introduction 1
- 1.1 Motivation . . . 1
- 1.2 Structure of the report . . . 1
- 1.3 Main contributions . . . 1

2 Background 3
- 2.1 Reinforcement Learning . . . 3
  - 2.1.1 Markov Decision Processes . . . 3
  - 2.1.2 Q-learning . . . 3
  - 2.1.3 Exploration/Exploitation . . . 3
  - 2.1.4 Deep Reinforcement Learning . . . 3
    - DQGAN . . . 3
  - 2.1.5 Problems with reinforcement learning techniques . . . 3
  - 2.1.6 Actor critic . . . 3
- 2.2 Generative Adversarial Networks . . . 3
  - 2.2.1 Architecture of GANs . . . 3
  - 2.2.2 Successes . . . 3
    - DCGAN . . . 3
  - 2.2.3 Conditional GANs . . . 3
- 2.3 Existing related work . . . 3
  - 2.3.1 Generative Adversarial Imitation Learning . . . 3

3 Environment 5
- 3.1 OpenAI Gym . . . 5
  - 3.1.1 Motivation . . . 5
  - 3.1.2 Algorithmic environments . . . 5
  - 3.1.3 MuJoCo and physics environments . . . 5
  - 3.1.4 Other environments . . . 5
- 3.2 Baseline: `FrozenLake-v0` . . . 5
  - 3.2.1 Description of the task . . . 5
  - 3.2.2 Motivation . . . 5
  - 3.2.3 Shortcomings . . . 5
- 3.3 Extended baseline: Randomised Frozen Lake . . . 5

4 Dataset creation 7
- 4.1 MapReduce job . . . 7

5 Adversarial Networks Training 9

6 Generative Adversarial Q-learning **11**
6.1 Using the trained Generator 11
6.1.1 Initialisation 11
6.1.2 Exploration 11
6.2 Using the trained Discriminator 11
6.2.1 Speeding up Q-learning on unseen maps 11

7 Scaling up to more complex tasks **13**
7.1 Larger Frozen Lakes 13
7.2 Physics environments 13

8 Conclusion **15**

Chapter 1

Introduction

1.1 Motivation

1.2 Structure of the report

1.3 Main contributions

Chapter 2

Background

2.1 Reinforcement Learning

2.1.1 Markov Decision Processes

2.1.2 Q-learning

2.1.3 Exploration/Exploitation

2.1.4 Deep Reinforcement Learning

DQGAN

2.1.5 Problems with reinforcement learning techniques

2.1.6 Actor critic

2.2 Generative Adversarial Networks

2.2.1 Architecture of GANs

2.2.2 Successes

DCGAN

2.2.3 Conditional GANs

2.3 Existing related work

2.3.1 Generative Adversarial Imitation Learning

Chapter 3

Environment

3.1 OpenAI Gym

3.1.1 Motivation

3.1.2 Algorithmic environments

3.1.3 MuJoCo and physics environments

3.1.4 Other environments

3.2 Baseline: `FrozenLake-v0`

3.2.1 Description of the task

3.2.2 Motivation

3.2.3 Shortcomings

3.3 Extended baseline: Randomised Frozen Lake

Chapter 4

Dataset creation

4.1 MapReduce job

Chapter 5

Adversarial Networks Training

Chapter 6

Generative Adversarial Q-learning

6.1 Using the trained Generator

6.1.1 Initialisation

6.1.2 Exploration

6.2 Using the trained Discriminator

6.2.1 Speeding up Q-learning on unseen maps

Chapter 7

Scaling up to more complex tasks

7.1 Larger Frozen Lakes

7.2 Physics environments

Chapter 8

Conclusion