Bipedal Walking - Reinforcement Learning

Roberto Figueiredo

April 2022

Abstract

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Acknowledgement

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Contents

Chapter 1

Introduction

In this chapter the target problem will firstly be introduced, as will a proposed solution to solve it. To understand the problem, background information will be covered to better understand the problem and what it is meant to target.

1.1 Robocup

Robocup is [3]

1.2 Humanoid Soccer League

1.3 Introduction to the Problem

1.4 Proposed Solution

Robotic locomotion has, until a few years ago, been focused on wheel-based movement. Although it is very stable and easy to implement, it lacks flexibility, the ability to move on uneven, unpredictable terrain and overcome obstacles such as stairs.

As a RoboCup team Bold Hearts member, which competes in the humanoid soccer league, our robots must walk, a recurring problem due to rule changes. As the objective of RoboCup is to achieve a realistic environment, competition rules change regularly. Changes in rules involve field changes, such as moving from flat ground to synthetic grass, enforcing that teams develop walking algorithms that can adapt to more variable environments. Rule changes also affect the robots, including their height, types of sensors and others. Changes in the robot's structure lead to the need to readapt the walking algorithms as they are dependent on these variables. These changes are time-consuming, and walking algorithms are a complex task requiring much effort from the team.

Chapter 2

Background Research

2.1 Learning Algorithms

2.2 Training Framework

2.3 Previous Implementations

2.4 Logging and Reproducibility

Chapter 3

Development Structure

Due to the complexity of the project a development structure has been put in place, this includes multiple steps of increasing complexity and realism, as the increasing complexity allowed for detecting problems at an earlier, simpler stage making the transition easier.

3.1 Structure

3.1.1 Cartpole

Cartpole is a classic exercise of reinforcement learning, it consists in balancing a pole in a cart moving on an horizontal plane, this environment allows for 2 discrete actions, which consist of applying force on the right or left side of the cart making it move in the oposite direction.

3.1.2 2D Walker

In this step a 2D environment of a simplified humanoid was used in order to train a walking behaviour, this new environment introduced a lot of new variables such as controlling multiple joints in a step, understanding how to efficiently calculate the best action and which algorithms to use. New challenges such as implementing a custom reward system, rendering and step functions where an important step in order to transition to 3D simulation.

3.1.3 3D Walker

3d simulation brings new challenges, such as a larger range of motion and more joints to controll, along with a more complex environment, requiring more processing power and more time to solve the problem. Along with this it requires a more complex reward system as a new dimention poses new problems.

3.2 Environment Definition

3.2.1 Cartpole

Its observation space consists of position of the cart on the horizontal axis and its velocity and the angle of the pole and its angular velocity. The objective of this environment is to balance the pole over 500 episodes To balance the pole the angle needs to stay in between $\pm 12°$ and the cart position stay in between the bounds of ± 2.4 The reward system is for cartpole is very simple, it earns 1 point for each time step survived. [4]

3.2.2 2D Walker

The 2d environment uses as a physics engine Pymunk [6], a python implementation of Chipmunk[1] in conjunction with pygame [5] to render the simulation.

To achieve walking a 2D simplified humanoid was developed in this environment, it consists of 8 joints, shoulder, hips, knees and ankle.

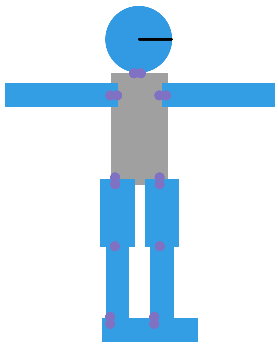

Figure 3.1: Representation of 2D humanoid

The reward system developed for this environment:

- Moves back: penalty of 200 points
- Stays in place: penalty of 100 points
- Moves forward: receives 0 points
- Both feet lose contact with the ground: cumulative penalty of 50 points
- Reaches target position: reward of 100 points
- Falls: penalty calculated as $\frac{1}{1-\gamma}\cdot$(higher penalty)

3.2.3 3D Walker

[2]

Chapter 4

Results

4.1 Cartpole Outcomes

4.2 2D Environment Outcomes

4.3 3D Environment Outcomes

4.4 Reward Function

Chapter 5

Future Research

5.1 Empowerment

5.2 Reward Function Development

5.3 Policy Gradients

5.4 Mujoco Implementation

5.5 Real Robot Training

Chapter 6

Project Evaluation

Chapter 7

Conclusion

Bibliography

[1] Chipmunk. Chipmunk. http://chipmunk-physics.net.

[2] Alberto Ezquerro, Miguel Angel Rodriguez, and Ricardo Tellez. openai ros. http://wiki.ros.org/openai_ros, 2016.

[3] Hiroaki Kitano, Minoru Asada, Yasuo Kuniyoshi, Itsuki Noda, and Eiichi Osawa. Robocup: The robot world cup initiative, 1995.

[4] OpenAI. Cartpole. https://github.com/openai/gym/blob/master/gym/envs/classic_control/cartpole.py, 2016.

[5] Pygame. Pygame. https://www.pygame.org.

[6] Pymunk. Pymunk. http://www.pymunk.org.