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 an attempt to advance the field of robotics by providing a common problem and an environment for sharing knowledge and collaborate. The first Robocup took place in 1997 in Nagoya, Japan. The Objective of Robocup is to achieve fully autonomous soccer playing robots that are able to defeat the FIFA World Cup champions by 2050.

Robocup has since evolved from just soccer and now includes multiple fields, Rescue, Soccer, @Home, Industrial and Junior. [3]

1.2 Soccer League

Robocup Soccer is split into multiple leagues, each with different challenges and focuses. The Small Size league uses small wheeled robots, each team is composed of six robots and play using a orange golf ball while tracked by a top-view camera; This enables the robots to abstract from challenges such as complex vision detection, walking and others, enabling the teams to focus on strategy and multi-robot/agent cooperation and control in a highly dynamic environment with a hybrid centralized/distributed system.

Middle size League assimilates to small size league, but in this case, the robots are of a larger size and must have all sensors on board; The main focus of the league is on mechatronics design, control and multi-agent cooperation at plan and perception levels.

Robocup also has simulation for most of its leagues, allowing the teams to focus on software and avoid the problems originated by using real robots hardware.

The Standard Platform is a step-up from the simulation league as while it uses real robots, NAOs, it allows the teams to focus mainly on software while using real robots and the challenges of using real robots without having to develop custom hardware. Each robot is fully autonomous and takes its one decisions.

The Humanoid league, assimilates the most to humans, using robots assimilating its shape and unlike humanoid robots outside the Humanoid League, the task of perception and world modeling is not simplified by using non-human like range sensors, making this, the most transversal league, requiring hardware and software development. In addition to soccer competitions technical challenges take place. Dynamic walking, running, and kicking the ball while maintaining balance, visual perception of the ball, other players, and the field, self-localization, and team play are among the many research issues investigated in the Humanoid League.

1.3 Bold Hearts

1.4 Introduction to the Project

1.4.1 Problem

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 member of the Bold Hearts team, which competes in the Teen size league, a branch of the Humanoid League, robots must walk and, in addition to the high complexity of the challenges imposed by the humanoid leagues, Robocup rules periodically change, these changes are put in place as the Robocup objective is to achieve the most realistic environment. Rule changes affect both robots, changing the required height, sensors and others, as well as affecting the environment such as moving from flat ground to synthetic grass.

One of the main challenges faced by the team is walking, which is one of the most complex movements performed by humans, requiring simultaneous control of multiple joints to move while maintaining balance. This is one of the most energy and time-consuming problems faced by the team addi-

tionally, the rules changes and continuous improvements require updates to the robot's structural architecture, leading to new updates to the walking algorithm.

1.4.2 Proposed Solution

Walking algorithms can be developed using various techniques, including explicit programming, supervised learning and unsupervised learning. Walking is very complex as there are a lot of variables involved in it, such as the ground contact, maintaining balance and the complex gait movement. The two main aspects important to highlight are that the walking algorithm requires changes when both the robot or external factors change and that it requires manual work from the team to achieve this.

The best way to solve the first mentioned problem requires having a robot and environment agnostic approach, this is possible by using Reinforcement Learning as the principles of the movement maintains therefore it should be possible to develop an agnostic reward function. From the moment a reinforcement learning solution is successfully implemented, the team should be able to use the same implementation to retrain the policy using the updated robot/environment that has been modeled in the simulator. This also allows to reduce the complexity of the problem as the input into the system is sensory data and the output is a set of actions to execute on the joints.

1.4.3 Aims and Objectives

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, 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.

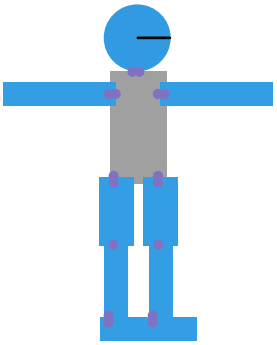

Figure 3.1: Representation of 2D humanoid

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.