Biped robot walking –

Reinforcement Learning

Initial Project Proposal

School of Computer Science

Roberto Figueiredo
SRN:17054162

Table of Contents

Introduction

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.

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. Due to the complexity of the problem, Reinforcement Learning is a good candidate. Along with being the best option due to the complexity, it is best to adapt to future modifications as the reward function is not based on any of the variables affected by rule changes.

To develop the reinforcement learning algorithm, two algorithms will be tested for the best performance, DeepQNetworks and Policy Gradient, chosen for the extensive documentation, expected performance, and straightforward implementation. The implementation will be using the Tensorflow library and Keras.

A combination of OpenAI Gym and Mujoco will be used to train the model. OpenAI Gym is a toolkit helpful in creating a training environment (Brockman, et al., 2016), standardizing and simplifying the implementation and result comparison. Mujoco is the simulator and physics engine recently open-sourced (Tassa, Y. et al., 2021). The reasoning

for this decision over the simulator currently used by the team, gazebo, is due to the excellent contact simulation, the future potential after the acquisition by the company DeepMind and the more straightforward integration with OpenAI Gym, with extensive examples and documentation. Gazebo will, although, be tested and compared alongside Mujoco to test for performance and implementation with OpenAI Gym as Gazebo has direct integration with ROS, not requiring a different control interface for the robot.

Supervisor Meeting Discussion

During the meeting discussion with the Supervisor, Daniel Polani, the main discussed topic was the solution to the proposed problem. The algorithms discussed were the two main algorithms mentioned, DeepQNetworks and Policy Gradients (Openai, 2021). While initially planned to use DeepQNetowrks due to the popularity and therefore a larger community along with better documentation and examples. During the discussion with the supervisor, it was decided to test both algorithms instead to attest the best performing yielding the most long-lasting results.

The supervisor proposed the simulator Mujoco due to being recently open-sourced and the excellent contact simulation. However, after reading a comparison between Gazebo and Mujoco, (Körber, M., et al., 2021). (before the DeepMind acquisition), Gazebo was the best performing of both, and the ROS integration supported the decision. The solution discussed was to have models for both simulators and compare the performance of this application on both simulators.

Preparatory Work

Tensorflow and keras

Tensorflow is a machine learning library tightly integrated with Keras, a neural network library (Tensorflow, 2021). Keras will be used to create and train the neural networks. While Keras is a straightforward tool, further preparatory work will be required to understand the underlying variables affecting the results, such as the different activation functions and layer types.

OpenAI Gym

OpenAI Gym is a toolkit used for training reinforcement learning. The decision to use is due to being the current standard and facilitates the benchmark and comparison of the results. It is integration with the simulator Mujoco is also well documented. OpenAI gym is a simple, well documented, easy to use tool. The integration with Mujoco needs further exploration of the custom environment definition.

Mujoco

Mujoco is a simulator and physics engine recently acquired by DeepMind and open-sourced. It was chosen over Gazebo due to being easier to integrate with OpenAI Gym, future potential, and new insights to the team. It will require further development of the control interface since ROS is not integrated.

Jupyter Notebook

To better document and visualize the development, It was decided to use Jupyter. This tool is helpful to split and organize the code for better interpretability by external observers.

Python

Python is the most widely used language in the machine learning and reinforcement learning field (Codecademy News. 2022) and, therefore, having the largest community, support and libraries, including OpenAI Gym.

Hardware

The simulation training will run mainly on a Macbook Pro M1 pro, using GPU acceleration. Further training can be run on the Bold Hearts team server if it shows performance improvements. If the development reaches the final stage successfully, tests will be performed using the Bold Hearts robots.

Stakeholders

This project involves the use of Reinforcement Learning techniques to develop a humanoid walking algorithm, a common problem in the RoboCup competition.

Any walking algorithm researcher can be interested in the outcome of the project, including the comparison of algorithms, simulators and the pain points found during development. More specifically, the RoboCup community where research of biped walking is prevalent, and since the competition is based on the sharing of knowledge and collaboration, other teams might benefit from the findings.

The principal stakeholder of the project is the Bold Hearts team, as the project is being developed in proximity and targeted at the needs, specifications, and robots of the team. Bold Hearts will benefit from the documentation being developed and the comparison of Mujoco with the current simulator, Gazebo. The continuation of the project will also be maintained by the team.

Risks

Risks during implementation:

Risk	Likelihood	Mitigation
Not having the required time to implement all variables of the project	Medium	Adapt requirements and reduce redundant variables such as different simulators.
Difficulty implementing a control interface	Medium	Access using Gazebo for last stage of the project.
Difficulty in implementing robot model in Mujoco	Low	Access using Gazebo for last stage of the project.

Risks due to implementation:

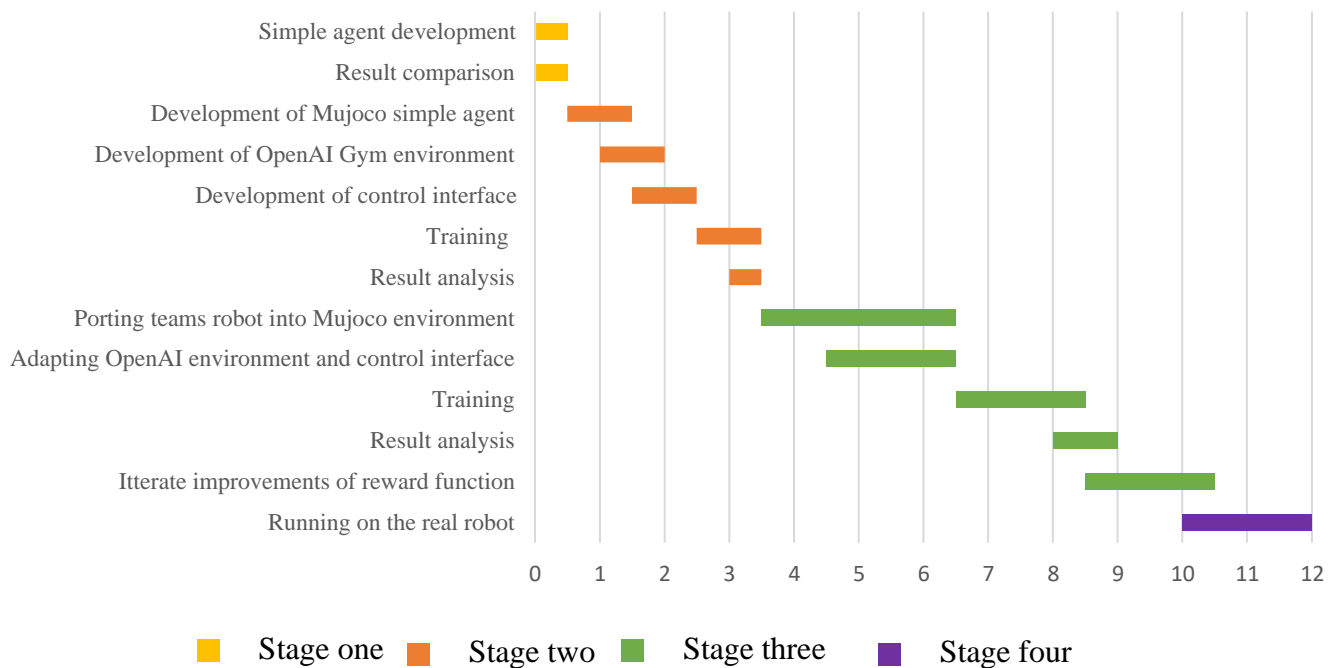| Risk	Likelihood	Mitigation
Reward function might require improvements	Medium	Adapt requirements and reduce redundant variables such as different simulators.
Choosing Mujoco might slow implementation into real robot	Medium	Access using Gazebo for last stage of the project.
Transition to the real robot isn't successful	Medium-High	Identification of disparities between real world behavior and simulation, iterate with implemented changes on simulator model or reward function.

Statement of Success

The project being developed covers three main topics, robotic biped locomotion, Reinforcement learning algorithm, simulator and physics engine. A successful conclusion of the project should output comprehensive documentation and knowledge on the technologies and processes mentioned.

Reinforcement learning includes many underlying technologies which the project will cover including the Markov Decision Process, Neural networks, activation functions and variables such as learning rate and discount factor. More in depth knowledge should be developed in the two selected algorithms, DeepQNetworks and Policy gradients, the project should output a comparison between the two, including the implementation process and the performance and stability of both, there should be a good understanding of the underlying working of the algorithms. The technologies used to implement the algorithms, Tensorflow and Keras, are in the core of the development therefore understanding how to implement them is essential for the success of the project.

The training of the algorithm will be implemented using OpenAI Gym and Mujoco. To successfully run training using OpenAI Gym it is important to understand the creation of environments, understanding the definition of state and action spaces and implementing a reward function. Using Mujoco will require modeling the robot and environment for the simulator, given that the robot is modeled on gazebo, understanding how to port the model into Mujoco is vital. A new control interface for the robot will be necessary as it does not integrate with ROS. A comparison between the Mujoco and Gazebo, simulators will be developed, to understand the benefits of each and the difficulty of implementing with OpenAI Gym.

Understanding the walking motion and how it applies to robots is essential for this project, mainly for the development of the reward function, to implement the reward function it is important to understand what movements and actions by the robot will lead to successful learning of walking.

Project Plan

Task	Stage
Simple agent development	Stage one
Result comparison	Stage one
Development of Mujoco simple agent	Stage two
Development of OpenAI Gym environment	Stage two
Development of control interface	Stage two
Training	Stage two
Result analysis	Stage two
Porting teams robot into Mujoco environment	Stage three
Adapting OpenAI environment and control interface	Stage three
Training	Stage three
Result analysis	Stage three
Itterate improvements of reward function	Stage three
Running on the real robot	Stage four

Legend: ■ Stage one ■ Stage two ■ Stage three ■ Stage four

Ethic

The information for the study in question is sourced from various origins including, software documentation, research papers of previous experiments and its conclusions and practical experimentation. The development doesn't require third parties to be involved as the main stakeholders are the Bold Hearts team belonging to the University of Hertfordshire. Since the project doesn't require input from external sources no ethics approval is required.

References

Körber, M. et al., 2021. Comparing Popular Simulation Environments in the Scope of Robotics and Reinforcement Learning. [online]
Available at: https://arxiv.org/pdf/2103.04616.pdf
[Accessed 3 January 2022]

Brockman, G. et al., 2016. *OpenAI Gym.* [Online]
Available at: https://arxiv.org/pdf/1606.01540.pdf
[Accessed 1 November 2021].

Codecademy News. 2022. *7 Top Machine Learning Programming Languages.* [online] Available at:
<https://www.codecademy.com/resources/blog/machine-learning-programming-languages/>
[Accessed 10 January 2022].

Tassa, Y. et al., 2021. Opening up a physics simulator for robotics. [Online]
Available at: https://deepmind.com/blog/announcements/mujoco
[Accessed 1 November 2021].

Tensorflow, 2021. *Why TensorFlow.* [Online]
Available at: https://www.tensorflow.org/about>
[Accessed 9 November 2021].

Openai, 2021. *Kinds of RL Algorithms — Spinning Up documentation..* [Online]
Available at: https://spinningup.openai.com/en/latest/spinningup/rl_intro2.html
[Accessed 14 November 2021].

Thompson, A. & George, N., 2016. *Deep Q-Learning for Humanoid Walking.* [Online] Available at:
https://web.wpi.edu/Pubs/E-project/Available/E-project-042616- 142036/unrestricted/Deep_Q-Learning_for_Humanoid_Walking.pdf
[Accessed 13 November 2021].

Wang, M., 2020. *Deep Q-Learning Tutorial: minDQN.* [Online]
Available at: https://towardsdatascience.com/deep-q-learning-tutorial-mindqn- 2a4c855abffc
[Accessed 2021 November 2021].

Hu, J. & Wellman, 2003. *Nash Q-Learning for General-Sum Stochastic Games.* [Online] Available at:
https://www.jmlr.org/papers/volume4/temp/hu03a.pdf
[Accessed 9 November 2021].

Huang, Q. et al., 2001. *Planning Walking Patterns for a Biped Robot.* [Online]
Available at: https://www.cs.cmu.edu/~cga/legs/kuff1a.pdf
[Accessed 1 November 2021].

Heess, N. et al., 2017. *Emergence of Locomotion Behaviours in Rich Environments.* [Online]
Available at: https://arxiv.org/abs/1707.02286
[Accessed 18 November 2021].
Haarnoia, T. et al., 2019. *Learning to Walk via Deep Reinforcement Learning.* [Online] Available at:
https://arxiv.org/pdf/1812.11103.pdf
[Accessed 1 November 2021].

Gong, D. & Zuo, G., 2010. *A Review of Gait Optimization Based on Evolutionary Computation.* [Online]
Available at:

https://www.researchgate.net/publication/220449058_A_Review_of_Gait_Optimization_B
ased_on_Evolutionary_Computation
[Accessed 1 November 2021].