Final Year Project Report

Full Unit - Project Plan

Resourceful Robots

Cougar Tasker

A report submitted in part fulfilment of the degree of

MSci (Hons) in Computer Science (Artificial Intelligence)

Supervisor: Dr. Anand Subramoney

Department of Computer Science
Royal Holloway, University of London

October 5, 2023

Declaration

This report has been prepared on the basis of my own work. Where other published and unpublished source materials have been used, these have been acknowledged.

Word Count:

Student Name: Cougar Tasker

Date of Submission: 05/10/2023

Signature: *Cougar Tasker*

Table of Contents

Abstract

Autonomous robots such as Boston-dynamic's Spot are increasing in prevalence across a wide range of fields[1]. Furthermore, these robots are increasingly integral in modern society, taking on ever more advanced tasks[2]. As autonomous robots take on increasingly complex and resource-intensive tasks, optimising their resource consumption while meeting objectives becomes an increasingly significant challenge. These autonomous robots operate in diverse environments with varying objectives, so a singular algorithm will not perform optimally for all cases.

This project aims to explore how reinforcement learning can provide a solution for effectively prioritising these objectives and managing resources. Reinforcement learning, a form of machine learning, involves an agent that can perceive and perform actions in an environment, learning to make optimal decisions[3]. Its ability to handle delayed rewards sets it apart from other machine-learning approaches[3], making it particularly well-suited for addressing this problem. In a similar problem, reinforcement learning has already proven to improve the efficiency of hybrid tracked vehicles over traditional dynamic-programming strategies[4]. This project differs from the work by Yuan Zou et al.[4] in several meaningful ways: The agent's perception will be the environment around the machine rather than its internals, and its actions and goals will be more abstract, such as positioning the vehicle to achieve objectives. This project will simulate the environment, initially using a grid world and potentially incorporating more advanced environments from OpenAI's gym library[5].

Reinforcement learning differs from supervised learning. In supervised learning, a model receives input and expected output data, usually supplied by a human[6]. In contrast, in reinforcement learning, the agent isn't given predefined correct actions; instead, it has to learn the right actions from exploring the environment[6]. This characteristic makes reinforcement learning well-suited for our problem since our agent can adapt to novel environments, just like in the real world, without needing external guidance. However, our agent will need a reward signal. For our problem, the definition is straightforward: the agent is rewarded for gathering resources and penalised for running out of energy. Importantly, our agent will perceive its energy capacity and the environment.

Motivations

My inspiration for studying this degree in artificial intelligence comes in large part from my belief that AI is becoming increasingly pivotal in shaping the future of technology and industry. For this purpose, this project presents an invaluable opportunity for my personal and professional growth. It is a fantastic platform to improve my comprehension of reinforcement learning while offering hands-on experience.

What is unique about this resource-gathering robot project is its structured progression of complexity, Starting from fundamental concepts and culminating in advanced techniques. This gradient makes the complex nature of reinforcement learning more approachable than it may be in industry.

This project interests me because of its generality and applicability to many different scenarios. Resource-gathering has the potential to incorporate many real-world constraints like energy, visibility and obstacles. I would like to see how this impacts different exploration strategies.

Last year, I completed my year-long internship at Zing Dev (Zing), a digital communications company that is progressively incorporating AI systems for its customers. This experience has demonstrated to me the value of understanding the internals of these AI systems. It is clear that AI is a clear focus for most companies, Ransbotham et al. said:

> Almost 85% believe AI will allow their companies to obtain or sustain a competitive advantage [7]

Through this project, I aim to improve my understanding of autonomous agents' benefits, biases, and limitations. This knowledge will be desirable for many companies like Zing working with artificial agents.

Objectives

The primary objective is to develop and understand reinforcement learning agents. The project will include two parts: a report and a graphical program. In the report, I will describe the reinforcement learning concepts that are implemented by the program, such as Markov decision processes and finding optimal policies by the Bellman equations. The report will also cover the software engineering process of the program and the results of different parameter choices and exploration strategies. The program should achieve the following goals:

- Energy-Efficient Navigation: Create a robot that can autonomously navigate a grid world while making optimal decisions to conserve energy. The robot should learn to prioritise energy-efficient paths.
- Resource Gathering Strategies: Design intelligent resource-gathering strategies for the robot, ensuring it collects essential resources while balancing energy expenditure. The robot should adapt its behaviour based on the availability of resources and its current energy levels.
- Dynamic Energy Management: The robot should monitor its energy reserves and adjust its actions accordingly. This includes learning when to engage in resource gathering and when to return to a charging station.
- Effective Exploration: Develop exploration strategies that enable the robot to learn and adapt to different grid world scenarios.
- Deep Reinforcement Learning: As an extension, the program should integrate deep neural networks (DNN) with Q-learning. The advantage of using a DNN instead of a table is the DNN's ability to generalise knowledge[8]. This unlocks working in environments with far more extensive observations and action spaces, allowing for added complexity to our existing environment or an extension of integrating a different environment, such as one from OpenAI's gym library[5].

Technology

The program will be developed in Python[9]; this is motivated by Python's strong ecosystem of tooling and community for data processing tasks. Performance is essential with machine learning tasks[10]. Python enables the use of low-level libraries that can use the hardware efficiently with high-level APIs[10]. These high-level APIs abstract over unimportant details, increasing productivity[10]. However, Raschka et al. states:

> Unfortunately, the most widely used implementation of the Python compiler and interpreter, CPython, executes CPU-bound code in a single thread, and its multiprocessing packages come with other significant performance trade-offs.[10]

.

For this reason, The project will start using PyPy. PyPy is a just-in-time (JIT) compiler that runs code four times faster according to PyPy's benchmarks[11][10]. Nevertheless, if the project runs into compatibility issues, the project may fall back to CPython.

To focus the program's development on its primary objectives, the application will use the libraries Kivy[12] and PyTorch[13]. Kivy is a library that provides functionality to create natural user interfaces quickly[14]. PyTorch is a Python library that provides the functionality for training Deep Neural networks with hardware (GPU) acceleration[15]. PyTorch will enable this application to integrate deep learning with the Q-Learning agent without getting caught up in implementing efficient tensor algorithms such as for backpropagation [15]. The application may also use other utility libraries, such as NumPy[16] for efficient calculations or Pandas[17] for data analysis.

Chapter 1: Timeline

1.1 Term One

Week	Goals	Explanation/Motivation
18/09/2023-22/09/2023	Study Reinforcement Learning chapter in Machine Learning[3]	This book has established a foundational understanding of reinforcement learning, which will serve as the basis for creating this project plan.
25/09/2023-29/09/2023	Create first project plan draft	Creating a project plan draft enables my supervisor to review and guide the final report in the right direction.
02/10/2023-06/10/2023	Complete project plan and its formatting	Incorporate enhancements suggested by my supervisor, refine document formatting to enhance its visual appeal, and gain proficiency in LaTeX for future reports.
09/10/2023-13/10/2023	Research reinforcement learning: • Read chapters 1,3 and 4 of Reinforcement Learning An Introduction [6] • Create the first draft of Markov decision processes (MDPs) report • Create the first draft of policy and value function report	Starting with research on reinforcement learning provides a strong foundation for application development. Additionally, early report drafting allows for increased opportunities for feedback
16/10/2023-20/10/2023	Research reinforcement learning: • Read chapter 6 of Reinforcement Learning An Introduction [6] • Create the first draft of the Q-learning report • Create the first draft report about learning as incrementally optimising policy in an MDP	This research helps deepen my understanding of reinforcement learning, and the reports serve the dual purpose of advancing toward the Interim submission while testing my knowledge.

Week	Goals	Explanation/Motivation
23/10/2023-27/10/2023	Initialise project: • Configure development environment • Project setup and configuration	This setup will establish the essential groundwork for application development, ensuring the proper configuration of the environment and tools to facilitate the creation of a robust application while upholding high code standards and quality.
30/10/2023-03/11/2023	• Model simple state and reward function for grid world. • Provide basic visualisation for a grid world state. • Mock the AI with hard-coded actions to validate the visualisation	The primary objective for this week is to create a vertical slice of the application, enabling rapid feedback and issue identification before committing to full-scale development.
06/11/2023-10/11/2023	Implementation of value iteration algorithm	As a planning algorithm, implementing this algorithm will efficiently generate optimal policies for our environment and grid world, providing a benchmark for our Q-learning implementation and assisting in the debugging process.
13/11/2023-17/11/2023	Implementation of Q-learning	The implementation of Q-learning will be an important focal point for the presentation. Initiating this process at an early stage will allow plenty of time to ensure its completion.
20/11/2023-24/11/2023	Finalise the report	As the Interim report deadline approaches, this week is dedicated to finalising the report, identifying errors, and making improvements wherever possible.
27/11/2023-01/12/2023	• Prepare for the presentation • Improve project graphics, add controls • Submit the interim report	Improving the graphics will help make the main functionality clear and appealing, and adding controls will provide interactivity that will help the program come across well for the presentation.
04/12/2023-08/12/2023	give the presentation	

1.2 Term Two

Week	Goals	Explanation/Motivation
08/01/2024-12/01/2024	Make the first draft of the poster	As a precaution, given the absence of a specific deadline, I intend to start the poster at the start of the term.
15/01/2024-19/01/2024	Develop different exploration strategies	Implementing multiple exploration strategies is important for my final report, so I am prioritising it.
22/01/2024-26/01/2024	• Record data from different exploration strategies and parameter choices • Analyse the exploration strategies' effectiveness	After creating the strategies, it is the ideal time to get their data and analyse it because I can iterate on the strategies to provide more in-depth research. However, I will need to be careful not to introduce data snooping.
29/01/2024-02/02/2024	Write a report on the effect of different parameter choices and exploration strategies	Following the recent data analysis, the report's writing follows naturally and benefits from the insights gained.
05/02/2024-09/02/2024 and 12/02/2024-16/02/2024	Implement deep learning	While a two-week timeline for implementing deep learning may be ambitious, leveraging the PyTorch library will simplify the process[15]. The remaining tasks would include integrating the model with the existing codebase and fine-tuning the model
19/02/2024-23/02/2024	Write the report on the software engineering process involved in generating this program	With the majority of development now complete, this is a good week to reflect on the whole engineering process and report on it.
26/02/2024-01/03/2024	Integrate agent with different environments from OpenAI's gym library[5]	As a further extension, it would be valuable to assess how well the agent generalises to entirely novel environments. However, if there are time constraints, this can be omitted without compromising the core objectives.
04/03/2024-08/03/2024	Report on the data from deep learning and how the agent generalised to new environments	The results obtained from deep learning will help elevate the section on different learning strategies, offering an additional perspective from a different type of environment.
11/03/2024-15/03/2024 and 18/03/2024-22/03/2024	Improve and Finalise report	Given that the report accounts for 30% of the final grade, it is crucial to allocate time to enhance and refine it according to my supervisor's feedback.

Week	Goals	Explanation/Motivation
25/03/2024-29/03/2024	Submit report	

Chapter 2: Risks And Mitigations

2.1 Hardware Issues

Likelihood: Low
Importance: High

Hardware failures, such as a lost or malfunctioning laptop, could disrupt the project's progress. Project data should be regularly backed up externally to reduce this risk. The project will be stored under Git with a GitHub remote repository. Under this arrangement, frequent code pushes upstream can minimise potential data loss. The report will be written on the Google Docs platform, providing cloud backup and version control.

2.2 Time Management Issues

Likelihood: Moderate
Importance: Moderate

Underestimating the time required for tasks could result in missing milestones or goals. Perfect time estimates are impossible; however, tasks can be subdivided appropriately to avoid significant surprises. Spreading out tasks evenly and leaving some buffer time can assist in avoiding work being cut short.

2.3 Machine Learning Risks

Likelihood: Moderate
Importance: Low

Machine learning can be a slow and computationally expensive task; this risks slowing the development or failing to train an effective model. The Q-learning table's complexity and size can be gradually increased and dynamically controlled to avoid computational bottlenecks. Manual fine-tuning of the system can be timely. Instead, the hyperparameters will be optimised with grid search.

2.4 Software Development Challenges

Likelihood: High
Importance: Moderate

Developing a complete graphical application can be complex, and unexpected software bugs may arise. Modern software engineering principles can minimise bugs and improve software quality. That is why this project will have version control, test-driven development, documentation, and static code analysis.

2.5 GUI Development Challenges

Likelihood: Moderate
Importance: Low

Designing and implementing the Graphical User Interface may be more time-consuming or challenging than anticipated. This project will use GUI development libraries or frameworks that streamline the process. Start GUI development early in the project to allow for iterative improvements.

2.6 Understanding of Reinforcement Concepts

Likelihood: Moderate
Importance: High

Reinforcement learning has several abstract concepts, such as MDPs, Bellman equations, dynamic programming, or Q-learning, which may lead to incorrect implementations or interpretations. Referencing textbooks and research papers to check results can validate findings. Thoroughly studying reinforcement learning concepts can avoid misinterpretations.

2.7 Optimistic time estimates

Likelihood: Moderate
Importance: Moderate

While not setting pessimistic goals is key to creating a worthwhile project, setting optimistic goals may not be realistic, as there are likely unforeseen obstacles. To mitigate the risks of missing cascading deadlines, the project plan is ordered in importance to the core objectives. For example, in the second term, different exploration strategies are implemented at the beginning; however, integrating environments from OpenAI's gym library[5] is at the end as it can be removed without compromising on the main objectives.

Chapter 3: Literature Review

Robots conquer the world [turning point][1]	This article shows the growing usage of autonomous robots in both household and industrial settings, demonstrating the growing value of their efficient operation
Autonomous systems in anaesthesia: where do we stand in 2020? A narrative review[2]	This article drives the potential value of this project further by showing how important and complex the tasks that autonomous agents are performing; it also shows how reinforcement learning approaches can improve effectiveness over conventional "proportional-integral-derivative" controllers
Machine Learning[3]	This book contains a chapter (13) that provides a fantastic overview of reinforcement learning with a succinct description of Q-learning
Reinforcement learning-based real-time energy management for a hybrid tracked vehicle[4]	This article details using reinforcement learning for an energy optimisation problem; although there are many differences with this project, it has enough similarities to provide a proof of concept for this project to build upon
Reinforcement learning: An introduction[6]	This book provides a approachable but detailed explanation of reinforcement learning concepts and algorithms that will form the research foundation for this project.
Reshaping business with artificial intelligence: Closing the gap between ambition and action[7]	This article provides clear evidence for the business value that artificial intelligence can provide. To complete this project, I must develop my understanding of reinforcement learning and artificial intelligence. This article shows that "respondents overwhelmingly agree that AI will both require employees to learn new skills within the next five years and augment their existing skills." the understanding of artificial intelligence could be key to providing value as an employee.
An introduction to deep reinforcement learning[8]	
Machine learning in Python: Main developments and technology trends in data science, machine learning, and artificial intelligence[10]	
PyPy performance benchmarks[11]	
Kivy–a framework for rapid creation of innovative user interfaces[14]	
Pytorch: An imperative style, high-performance deep learning library[15]	

3.1 Documentation

I have cited the documentation for the technologies I intend to use, as this is the definitive source of information regarding their functionality and an excellent place to learn. These are the technologies mentioned:

- Gym Documentation[5]
- Python Documentation[9]
- Kivy Documentation[12]
- PyTorch Documentation[13]
- NumPy Documentation[16]
- Pandas Documentation[17]

Bibliography

[1] M. Hagele, "Robots conquer the world [turning point]," *IEEE Robotics & Automation Magazine*, vol. 23, no. 1, 2016.

[2] C. Zaouter, A. Joosten, J. Rinehart, M. M. Struys, and T. M. Hemmerling, "Autonomous systems in anesthesia: where do we stand in 2020? a narrative review," *Anesthesia & Analgesia*, vol. 130, no. 5, 2020.

[3] T. Mitchell, *Machine Learning*, ser. McGraw-Hill International Editions. McGraw-Hill, 1997.

[4] Y. Zou, T. Liu, D. Liu, and F. Sun, "Reinforcement learning-based real-time energy management for a hybrid tracked vehicle," *Applied Energy*, vol. 171, pp. 372–382, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261916304081

[5] OpenAI. Gym documentation. [Online]. Available: https://www.gymlibrary.dev

[6] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[7] S. Ransbotham, D. Kiron, P. Gerbert, and M. Reeves, "Reshaping business with artificial intelligence: Closing the gap between ambition and action," *MIT Sloan Management Review*, vol. 59, no. 1, 2017.

[8] V. François-Lavet, P. Henderson, R. Islam, M. G. Bellemare, J. Pineau *et al.*, "An introduction to deep reinforcement learning," *Foundations and Trends® in Machine Learning*, vol. 11, no. 3-4, 2018.

[9] P. S. Foundation. Python documentation. [Online]. Available: https://docs.python.org/3/

[10] S. Raschka, J. Patterson, and C. Nolet, "Machine learning in python: Main developments and technology trends in data science, machine learning, and artificial intelligence," *Information*, vol. 11, no. 4, p. 193, 2020.

[11] T. P. Team. Pypy performance benchmarks. [Online]. Available: https://speed.pypy.org/

[12] K. Organization. Kivy documentation. [Online]. Available: https://kivy.org/doc/stable

[13] P. Foundation. Pytorch documentation. [Online]. Available: https://pytorch.org/docs/stable/index.html

[14] M. Virbel, T. Hansen, and O. Lobunets, "Kivy–a framework for rapid creation of innovative user interfaces," in *Workshop-Proceedings der Tagung Mensch & Computer 2011. überMEDIEN/ ÜBERmorgen*. Universitätsverlag Chemnitz, 2011.

[15] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, 2019.

[16] Numpy documentation. [Online]. Available: https://numpy.org/doc/stable

[17] Pandas documentation. [Online]. Available: https://pandas.pydata.org/docs