Untitled Intrinsically Motivated Reinforcement Learning

Jordan de Burgo

November 4, 2022

Contents

1 Problem Description

1.1 What is Reinforcement Learning

Reinforcement learning is a technique in artificial intelligence that maps situations to actions to maximise a numerical signal reward, and the learner is not told which actions to take but instead must discover which actions yield the most reward by trying them (Sutton and Barto, 2018). Essentially it is a trial-and-error approach that adjusts a numerical reward based on the outcome in the environment. This approach aims to make an agent learn through interaction with an environment it is not familiar with to achieve long-term goals. A challenge that arises with this is the trade-off between looking at new actions and repeating actions that it has tried in the past that have resulted in a good reward. It is this value function that is key for developing an algorithm that will learn optimally.

1.2 Applications of Reinforcement Learning

There have been many breakthroughs in reinforcement learning, being able to play Atari games, AlphaZero, and DeepStack (Li, 2019). These achievements all represent environments where different data is available to the learner. Atari games are single-player games. AlphaZero represents two-player games where the learner has access to all the information about the environment. DeepStack represents a two-player game where not all of the information is known. There are many cases in which automating something in this kind of environment can be helpful. There are several situations in the healthcare industry where reinforcement learning could be applied to help solve problems; for example, reinforcement learning methods have been studied in deriving efficient treatment strategies for cancer chemotherapy. It was used in simulation to minimize tumor cell population and drug amount whilst maximizing population of normal and immune cells. It's aim was to discover an effective drug scheduling policy, and it's performance was shown to be better than traditional administration of the drug (Yu, Liu and Nemati, 2019). There is also a use in robotics; it is possible to train robots to be dexterous. These robots can even be trained within a simulated environment before deployment. This can minimise any damage done to objects in the real world when the robot is being trained, especially if the robot is handling fragile objects. There are also many other potential applications in energy, finance, transportation, and other industries.

1.3 What is Intrinsically Motivated Reinforcement Learning

Intrinsically motivated reinforcement learning is a potential solution to the challenge described earlier, the exploration problem. Most current reinforcement learning methods will choose states to visit based on extrinsically given rewards (rewards given by the environment), whereas an intrinsically motivated agent will choose states (at least partially) based on a reward it gives itself for how "novel" the state is. The novelty of a state is defined by how unique a state appears to the agent, if an agent has seen a state (or a similar state in some cases) lots of times the state would not be very novel and therefore would receive little intrinsic reward by the agent. In many real-world scenarios extrinsic rewards to the agent are extremely sparse, and it is not possible to construct a shaped reward function. This is a problem as a traditional agent can only update its policy when it reaches a goal state, and hoping to stumble into one of these goal states by random exploration is unlikely to work in these sparse reward scenarios (Pathak et al., 2017). Intrinsically motivated reinforcement learning helps to address this problem by providing a way for an agent to update its policy before reaching a goal state, and will help the agent to work its way towards a goal state, if one exists, systematically rather than randomly. It is in these sparse extrinsic reward environments that intrinsically motivated reinforcement learning should have the most benefit for application, for example, in developmental robotics.

1.4 Project Aim

To investigate intrinsically motivated reinforcement learning methods and implement one in different environments to evaluate it.

1.5 Project Objectives

- Understand and replicate a method of intrinsically motivated reinforcement learning and (some of) the existing results in the literature using this method. (I am currently leaning towards a curiosity-driven/prediction error based method).

- Evaluate on different test environments to assess the method's strengths and weaknesses.

- Suggest algorithmic improvements and attempt to implement these and evaluate their effectiveness.

2 Requirements Specification

2.1 General Objectives

1. **High Priority** To understand a method of intrinsically motivated reinforcement learning
2. **High Priority** To implement a method of intrinsically motivated reinforcement learning in very simple environments
3. **Medium Priority** To implement a method of intrinsically motivated reinforcement learning in more complex environments
4. **Low Priority** To suggest an improvement to the algorithm
5. **Low Priority** To implement and evaluate the suggested improvement(s)

2.2 Project Requirements

1. The project must be completed by the 5th of May 2023
2. The project must include a Literature and Technology Survey
3. A method of intrinsically motivated reinforcement learning must be implemented in a simple environment by the demonstration of progress on the 20th of February 2023
4. The project should stay reasonably on track with the project plan

3 Project Plan

3.1 Initial Gantt Chart

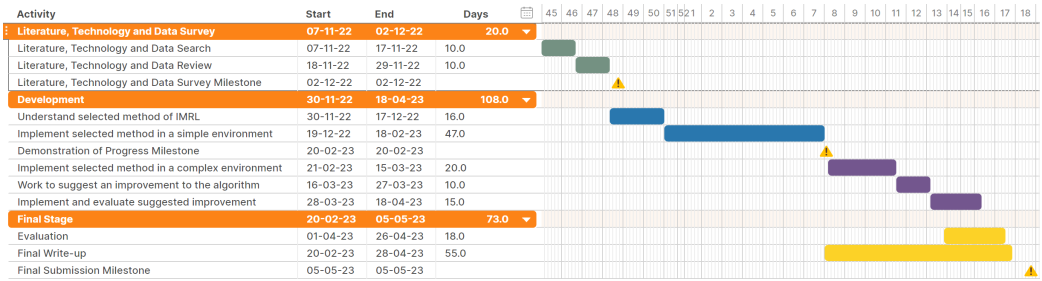

Figure 1: Gantt Chart outlining the initial plan

Task	Start Date	End Date	Effort (Hours)
Literature, Technology and Data Survey	07/11/22	29/11/22	40
Understand selected method of IMRL	30/11/22	17/12/22	50
Implement selected method in a simple environment	19/12/22	18/02/23	100
Implement selected method in a complex environment	21/02/23	15/03/23	40
Suggest, implement, and evaluate an improvement	16/03/23	18/04/23	50
Evaluation	01/04/23	26/04/23	30
Final Write-up	20/02/23	28/04/23	110

Table 1: Outline of tasks and their estimated duration

3.2 Tasks and Milestones

The 4 milestones presented and the deadlines are:

1. Project Proposal - 8PM Friday the 4th of November 2022
2. Literature and Technology Survey - 8PM Friday the 2nd of December 2022
3. Demonstration of Progress - 8PM Monday the 20th of February 2023 (TBC)
4. Dissertation - 8PM Friday the 5th of May 2023 (TBC)

The plan has been crafted such that I aim to get to all milestones and deadlines with time to review my work thoroughly and ensure that the impact of any setback is minimised, in the worst case low priority objectives may not be met. The number of hours (seen in Table 1) each task has been allocated is based on the recommended number of hours to spend on the project, with some additional hours to ensure that every task is completed to a high standard. My plan accounts for taking Sundays off and some time around Christmas, new years, and my birthday. This also allows me the plan of using Sundays if I have not gotten enough done during the week.

3.3 Risk Assessment

3.3.1 Illness

If I get ill, there should be enough hours within the working time frame to still complete the task if it is not a serious illness.

3.3.2 PC Failure

All work will be backed up on GoogleDrive and the Bath Files System. All code will be backed up on GitHub.

3.3.3 Tasks Overrun

All tasks finish with enough time before any deadlines to mitigate this risk.

3.3.4 Effort Underestimated for Tasks

Objectives have been sorted by priority such that even if not all objectives are met, I will have some working product to present.

4 Resources

4.1 Literature Resources

4.1.1 Research Papers related to Reinforcement Learning

These will be used to further my understanding of the topic as a whole, to investigate any research that may bring new ideas to my own project. I will use papers available to me through the university library and papers freely available online. The papers I currently plan to read around this topic are:

- Playing Atari with Deep Reinforcement Learning (Mnih et al., 2013)

- Offline Reinforcement Learning: Tutorial, Review, and Perspectives on Open Problems (Levine et al., 2020)

- A Survey on Offline Reinforcement Learning: Taxonomy, Review, and Open Problems (Prudencio, Maximo and Colombini, 2022)

4.1.2 Books on Reinforcement Learning

These will be used to gain further insight into the fundamentals of reinforcement learning, covering well-researched topic areas to expand my knowledge and aid understanding of how reinforcement learning algorithms are implemented. Books from the university library and from my own collection will be used. I am currently reading Reinforcement Learning: An Introduction (Sutton and Barto, 2018)

4.1.3 Research Papers related to Intrinsically Motivated Reinforcement Learning

These will be used to understand different methods of intrinsically motivated reinforcement learning and select papers will be used in my attempt to replicate the results of them. Papers in this area that I have read or currently plan to read are:

- Large-Scale Study of Curiosity-Driven Learning (Burda et al., 2018)

- Exploration by random network distillation (Burda et al., 2019)

- Curiosity-driven Exploration by Self-supervised Prediction (Pathak et al., 2017)

- Self-Supervised Exploration via Disagreement (Pathak, Gandhi and Gupta, 2019)

- Count-Based Exploration with Neural Density Models (Ostrovski et al., 2017)

- Curiosity-Bottleneck: Exploration by Distilling Task-Specific Novelty (Kim et al., 2019)

- A Possibility for Implementing Curiosity and Boredom in Model-Building Neural Controllers (Schmidhuber, 1991)

4.2 Computational Resources

4.2.1 OpenAI Gym

Gym is an open source Python library that provides an API that allows learning algorithms to communicate with environments. The environments chosen to test my algorithm(s) will be selected from environments within the Gym library.

4.2.2 Personal Computer

My own computer will be used to develop and train the algorithms for my project. It will also be used for the final write up of the project.

4.2.3 GPU Cluster

In the event that my PC fails, or should it not provide sufficient computational power, I may benefit from being able to access the department's GPU cluster to train my algorithms on as appropriate.

References

Burda, Y., Edwards, H., Pathak, D., Storkey, A., Darrell, T. and Efros, A.A., 2018. Large-scale study of curiosity-driven learning [Online]. Available from: https://doi.org/10.48550/ARXIV.1808.04355.

Burda, Y., Edwards, H., Storkey, A. and Klimov, O., 2019. Exploration by random network distillation. *International conference on learning representations* [Online]. Available from: https://openreview.net/forum?id=H1lJJnR5Ym.

Kim, Y., Nam, W., Kim, H., Kim, J.H. and Kim, G., 2019. Curiosity-bottleneck: Exploration by distilling task-specific novelty. In: K. Chaudhuri and R. Salakhutdinov, eds. *Proceedings of the 36th international conference on machine learning* [Online]. PMLR, *Proceedings of machine learning research*, vol. 97, pp.3379–3388. Available from: https://proceedings.mlr.press/v97/kim19c.html.

Levine, S., Kumar, A., Tucker, G. and Fu, J., 2020. Offline reinforcement learning: Tutorial, review, and perspectives on open problems [Online]. Available from: https://doi.org/10.48550/ARXIV.2005.01643.

Li, Y., 2019. Reinforcement learning applications. *Corr* [Online], abs/1908.06973. 1908.06973, Available from: http://arxiv.org/abs/1908.06973.

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D. and Riedmiller, M., 2013. Playing atari with deep reinforcement learning [Online]. Available from: https://doi.org/10.48550/ARXIV.1312.5602.

Ostrovski, G., Bellemare, M.G., Oord, A. van den and Munos, R., 2017. Count-based exploration with neural density models. In: D. Precup and Y.W. Teh, eds. *Proceedings of the 34th international conference on machine learning* [Online]. PMLR, *Proceedings of machine learning research*, vol. 70, pp.2721–2730. Available from: https://proceedings.mlr.press/v70/ostrovski17a.html.

Pathak, D., Agrawal, P., Efros, A.A. and Darrell, T., 2017. Curiosity-driven exploration by self-supervised prediction [Online]. Available from: https://doi.org/10.48550/ARXIV.1705.05363.

Pathak, D., Gandhi, D. and Gupta, A., 2019. Self-supervised exploration via disagreement. In: K. Chaudhuri and R. Salakhutdinov, eds.

Proceedings of the 36th international conference on machine learning [Online]. PMLR, *Proceedings of machine learning research*, vol. 97, pp.5062–5071. Available from: https://proceedings.mlr.press/v97/pathak19a.html.

Prudencio, R.F., Maximo, M.R.O.A. and Colombini, E.L., 2022. A survey on offline reinforcement learning: Taxonomy, review, and open problems [Online]. Available from: https://doi.org/10.48550/ARXIV.2203.01387.

Schmidhuber, J., 1991. A possibility for implementing curiosity and boredom in model-building neural controllers. In: J.A. Meyer and S.W. Wilson, eds. *Proc. of the international conference on simulation of adaptive behavior: From animals to animats* [Online]. MIT Press / Bradford Books, pp.222–227. Available from: https://mediatum.ub.tum.de/doc/814958/814958.pdf.

Sutton, R.S. and Barto, A.G., 2018. *Reinforcement learning: An introduction*. 2nd ed. The MIT Press.

Yu, C., Liu, J. and Nemati, S., 2019. Reinforcement learning in healthcare: A survey. *Corr* [Online], abs/1908.08796. 1908.08796, Available from: http://arxiv.org/abs/1908.08796.