Reinforcement Learning: Transfer Learning Between Different Games

Student Name: Matthew Chapman

Supervisor Name: Dr Lawrence Mitchell

Submitted as part of the degree of BSc Natural Sciences to the

Board of Examiners in the Department of Computer Sciences, Durham University

March 23, 2021

***Abstract* —**

Background — Reinforcement learning, concerned with how agents learn behaviour through trial-and-error interactions with a dynamic environment, has proven to be a successful technique of achieving at least human-level performance by machines. Transfer learning is the application of knowledge gained while solving one problem to solve a different but related problem, and is a technique that allows reinforcement learning agents to adapt to new environments.

Aims — The aim of this project is to investigate the usefulness of transfer learning in a modern reinforcement system, where environments are Atari video games. We seek to quantify the benefit of pre-training on a game on an agent's ability to learn to play another game with similar mechanics.

Method — The method is to implement a modern reinforcement learning algorithm and train it on two selected games. Once the agent is performing well in the environment, we save a video of it playing, and use the video as the basis of pre-training of another agent, which we then evaluate on the test game.

Results — Pre-training leads to the agent performing better, as measured by an improvement in the following three metrics: jump-start performance, accumulated reward, and final performance. The more pre-training information the agent has, the better it performs.

Conclusions — Transfer learning is a promising method to prepare agents for environments where it has limited opportunities to interact with and learn from. By training agents on similar environments, we can build confidence that the agent will perform successfully when evaluated on the test environment. This is particularly useful in fields such as autonomous driving, where the vehicle must be able to adapt to various changes in the environment.

***Keywords* —** Artificial intelligence, machine learning, reinforcement learning, deep learning, transfer learning.

I INTRODUCTION (2-3 PAGES)

This project is about reinforcement learning and transfer learning. The project involves developing a reinforcement learning algorithm to learn to perform successfully in some environment. The project also involves investigating how transfer learning lets the algorithm store knowledge and apply it — to learn to perform successfully in a different but related environment.

A Background

A.1 Reinforcement learning

Reinforcement learning is the class of problems concerned with an agent learning behaviour through trial-and-error interactions with a dynamic environment (Kaelbling et al. 1996). An example of a problem is an aspiring tightrope walker (the agent) learning to maintain balance (the behaviour) while walking along a tightrope that contorts and wobbles under their weight (the dynamic environment). With each attempt and fall (the trial-and-error interactions), the walker learns how better to correct their balance, and adjusts their behaviour slightly for the next attempt. When the walker is able to maintain balance consistently over consecutive attempts, the desired behaviour is achieved, and so the learning task is complete. We say that the problem is solved and the reinforcement learning agent has learned to perform successfully in the environment.

There are algorithms that act as agents that solve reinforcement learning problems. These reinforcement learning algorithms can solve problems in physical settings, such as driving cars, or in virtual settings, such as playing games. We can treat these algorithms as functions that takes as input observations and outputs actions. Examples of observations are the video from a camera attached to a self-driving car or the positions of pieces on a chessboard in an online match. Examples of corresponding actions are to turn the steering wheel in one direction or to move a chess piece. The goal of the algorithm is to learn which actions are the best to take given some observations. How good an action given an observation is can be measured by how likely taking the action is to lead to the desired behaviour. For an algorithm that drives cars, the desired behaviour might be to drive safely, and so the algorithm would know to stop at a red traffic light. Whereas, for an algorithm that plays chess, the desired behaviour might be to win, and so the algorithm would know to take the opponent's king. The algorithm is learning a mapping, from actions and observations to values, to inform its decision-making. This mapping is initially unknown, but improves the more the algorithm interacts with its environment — the same way one gets better with practice at driving or playing chess. For relatively complex problems, there may not be an optimal solution, such as behaviour that guarantees no accidents or that always wins, and so the best the algorithm can do is to approximate an optimal solution.

A.2 Transfer learning

Transfer learning is the application of knowledge gained while solving one problem to solve a different but related problem (Sammut & Webb 2010). An example is an agent who has learned to walk a tightrope (solving one problem) applying their balancing ability (the knowledge gained) to learn to surf (a different but related problem). By reusing knowledge from solving past problems, it is expected that solving a different but related problem will be more efficient than it would be without the prior knowledge. As in the example, a tightrope walker should learn to balance a surfboard more easily and more quickly than someone without the same acquired balancing ability.

In transfer learning for reinforcement learning algorithms, the knowledge gained that can be applied is the agent's policy. The policy is the set of rules that determine an agent's behaviour (which can be informed by the mapping mentioned in the previous section). An example of a policy for an agent playing the video game Breakout might be to move the paddle randomly. Another, better policy might be to move the paddle in the direction of the projectile, so as to

deflect it. The policy of interested, however, is one that the reinforcement learning algorithm developed itself while learning to play. Depending on the architecture of the algorithm, the policy could be a neural network, so that developing the policy equates to training the network. An example of transfer learning for reinforcement learning algorithms, then, could be to apply the trained neural network, or part of it, that was trained in the agent that learned to play Breakout, to a new agent that is learning to play a different but related game, such as Pong.

A.3 Context

Reinforcement learning has proven to be a successful technique of achieving at least human-level performance by machines. However, real-world reinforcement learning agents can often only interact with their environment a limited number of times, due to reasons such as cost and risk of damage. In order to build confidence that an agent will perform successfully in a real-world environment, the agent can be trained in different but related environments, such as virtual simulations, where there is no limit to the number of interactions. Transfer learning is an effective method to prepare agents for environments with which they will have little to no interaction with before testing or deployment.

B Aims and achievements

B.1 Aims

The aim of the project is as follows: *To investigate the usefulness of transfer learning between Atari games by quantifying the benefit of pre-training*. To address this aim, the objectives for the project were divided into three categories: minimum, intermediate, and advanced.

The minimum objectives were to train 2 good policies for 2 Atari games, and generate 10,000 trajectories of 1,000 steps each from the policy for each game.

The intermediate objectives were to fit a generative model to the trajectories produced by 1 of the games, and transfer the model to the 2nd game.

the advanced objectives were to train n good policies for n Atari games, and fit a generative model to the trajectories produced by n-1 of the games, then transfer that model to the nth game.

The research questions were, *How large does the model need to be for the pre-training to be useful?* and *How does the size of the effect change when the amount of data is reduced by 10x? By 100x?*.

B.2 Achievements

What was achieved is as follows:

- We developed a reinforcement learning algorithm that learned to perform successfully in some environment.
- We applied the trained algorithm to a different but related environment and measured its performance.

II RELATED WORK (2-3 PAGES)

[Bridging paragraph] Explain why you're covering specific techniques and their relevance ...

- Reinforcement learning ... [critical analysis: issues, lead to] ...
- Deep reinforcement learning ... [critical analysis: issues, lead to] ...
- Transfer learning ... [critical analysis: issues, lead to] ...
- This relates to my research question because ...
- Motivate why did things I did ...

III SOLUTION (4-7 PAGES)

[Bridging paragraph] Rather than implement everything from scratch, build on frameworks to allow focus on algorithm design ...

A Specification and design

- The design of the solution was ...
- The architecture and architectural diagram of the solution were as follows ...

B Implementation issues

- The features of the implementation process were ...

C Tools and algorithms used

- The tools used were ...
- The algorithms used were ...

D Verification and validation

- Verification was done by ...
  - Do implementations work there? ...
  - What do I do to judge the outcome/success?
  - Try to answer whether transfer learning is generalisable
- Validation was done by ...

E Testing

[Bridging paragraph]

- Testing was done by ...
  - reproduce on simple problems

IV RESULTS (2-3 PAGES)

[Bridging paragraph]

A Evaluation method

- The evaluation methods adopted were ...

B Experimental settings

- These were the experimental settings for each experiment carried out: ...

C Results

- The results generated by the software were ...

V EVALUATION (1-2 PAGES)

[Bridging paragraph]

A Suitability of the approach (more SE, maybe exclude?)

- The approach was/was not suitable because ...
- Was it a good idea to use PyTorch, etc.?

B Strengths and limitations of the algorithm

- The strengths of the algorithm were ...
- The limitations of the algorithm were ...
- The lessons learnt were ...

C Project organisation

- The project was organised as well as you would expect in a global pandemic ...

VI CONCLUSIONS (1 PAGE)

A Project overview

- The project was to ...

B Main findings

- The main findings were as follows: ...
- The conclusions from these findings were ...

C Further work

- The project can be extended by ...

()

References

Kaelbling, L. P., Littman, M. L. & Moore, A. W. (1996), 'Reinforcement learning: A survey', *Journal of Artificial Intelligence Research* **4**, 237–285.
URL: *https://doi.org/10.1613/jair.301*

Sammut, C. & Webb, G. I., eds (2010), *Encyclopedia of Machine Learning*, Springer US.
URL: *https://doi.org/10.1007/978-0-387-30164-8*