

# 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 22, 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 is an aspiring tightrope artist (the agent) learns to walk from one end of a rope to another without falling (the behaviour) by repeatedly correcting their balance (the trial-and-error interactions) whenever the rope wobbles beneath them (the dynamic environment). With each failed attempt, the agent learns, and eventually succeeds. When the agent succeeds consistently, we say it has learned to successfully operate in its environment.

Reinforcement learning algorithms are algorithms that act as agents that learn behaviour in virtual environments. An example is to obtain the highest score in an Atari game. The algorithm can be to take as input the pixels from some frames, which we call a state, and then output the best action, such as to move the paddle left. How good an action from a state is can be quantified by giving it a value, such as the predicted final score if you moved left, versus if you moved right. As the algorithm gets more experience playing, its predictions get better. It's learning a mapping, that's initially unknown, from states and actions to points, or rewards. This is called the value function, and must be approximated for complicated environments. The approach described is called Q-learning. There are many approaches (described in Section II), with modern approaches incorporating the use of deep neural networks.

### **A.2 Transfer learning**

Transfer learning is the application of knowledge gained while solving one problem to solve a different but related problem. An example is a successful tightrope artist applying their balancing skills to learn to surf. Learning to walk a tightrope is solving one problem, whereas learning to surf is solving a different but related problem. The knowledge gained, and then applied, are the balancing skills, which relates the two tasks. Similarly, learning tennis and table-tennis are two problems related by the knowledge of hitting a ball over a net with a racket.

In transfer learning for reinforcement learning, the knowledge is the agent's policy. The policy is the set of rules that determine an agent's behaviour. An example of a policy for an agent playing the Atari game Breakout is to press buttons randomly. Another, better policy is to press buttons so that the paddle moves in the direction of the projectile, so as to hit it. Given two different but related games, such as Breakout and Pong, it makes sense for a reinforcement learning algorithm to apply the policy gained while learning to play one game to learn to play the other.

### **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  $n$ th 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>