ABSTRACT

ANALYSIS OF Q- LEARNING BASED GAME PLAYING AGENTS FOR ABSTRACT BOARD GAMES WITH INCREASING STATE-SPACE COMPLEXITY

by Indrima Upadhyay

This thesis investigates Q-learning agents in a Reinforcement Learning framework for abstract board games. The two key contributions are: exploring the training of Q-learning agents, and a methodology to evaluate different agents playing abstract games. We focus on - Tic-Tac-Toe, Nine-Men's Morris, and Mancala, noting that each of these games is solved.

To train our Q-agent, we test the impact of a teaching agent (Q-learning twin, deterministic Min-Max, and a non-deterministic Min-Max) based on the number of training epochs needed until an agent converges. We find that a deterministic Min-Max is the best teaching agent, but a Q-learning twin allows us to converge without needing a pre-existing agent. In training, we include a methodology to create weaker agents for entertainment purposes.

To evaluate a range of competitive agents, we provide a methodology and conduct a round-robin tournament. We find that a deterministic Min-Max agent scores maximum points for all three games, the Q-learning based agent places second for both Tic-Tac-Toe and Nine Men's Morris, and a non-deterministic Min-Max places second for Mancala.

From these two contributions we summarize our conclusions and provide discussion on these results and provide insight on how this space might be further investigated.

ANALYSIS OF Q- LEARNING BASED GAME PLAYING AGENTS FOR ABSTRACT BOARD GAMES WITH INCREASING STATE-SPACE COMPLEXITY

A Thesis

Submitted to the

Faculty of Miami University

in partial fulfillment of

the requirements for the degree of

Master of Science

by

Indrima Upadhyay

Miami University

Oxford, Ohio

2021

Advisor: Dr. Peter Jamieson

Reader: Dr. Chi-Hao Cheng

Reader: Dr. Ran Zhang

This Thesis titled

ANALYSIS OF Q- LEARNING BASED GAME PLAYING AGENTS FOR ABSTRACT BOARD GAMES WITH INCREASING STATE-SPACE COMPLEXITY

by

Indrima Upadhyay

has been approved for publication by

The College of Engineering and Computing

and

The Department of Electrical & Computer Engineering

Dr. Peter Jamieson

Dr. Chi-Hao Cheng

Dr. Ran Zhang

Table of Contents

List of Tables

List of Figures

Dedication

To my parents, Mrs. Nandita Upadhyay, and Dr. Ajay Kumar Upadhyay. Thank you for always being my best friends, my support system and everything else in between.

To my sister, Dr. Ipshita Upadhyay and brother-in-law, Mr. Shashank Bajpai. Thank you for showing me nothing is unachievable and for always being there whenever I needed.

Acknowledgements

Writing a thesis is harder than I thought and more rewarding than I could have ever imagined. Throughout the writing of this thesis I have received a great deal of support and assistance.

This thesis and the research behind it would not have been possible without the remarkable guidance of my advisor, Dr. Peter Jamieson, Associate Professor in the Department of Electrical and Computer Engineering, Miami University. His supervision, advice, guidance from the very early stage of this research and timely help took me towards the completion of this thesis. His continuous encouragement and support not only aided me in completing the thesis but also inspired me to explore my career in the pertinent industry. It is indeed a great pleasure to work under the guidance of Dr. Peter Jamieson. His knowledge and thorough attention to detail has been an absolute inspiration.

I am also deeply indebted to Mr. Michael Bomholt for constantly providing his insight and expertise that greatly assisted the research by proving to be indispensable support and advice. His patience, motivation, enthusiasm, and immense knowledge helped me throughout the course of this research and writing of thesis.

I am very thankful to my thesis committee members Dr. Chi Hao Cheng and Dr. Ran Zhang for their constructive criticism and feedback that helped to improve my thesis. I am grateful to Miami University for providing me with ample opportunities and the perfect environment to grow not only as a research student but also as an individual. My fellow graduate students and friends in the Electrical and Computer Engineering Department, Miami University made the entire experience all the more enriching and enjoyable. Thanks a lot for all the stimulating and interesting discussions we have had over the last two years

It would be remiss if I don't thank my parents, Dr. Ajay Kumar Upadhyay and Mrs. Nandita Upadhyay, for their unwavering faith in me and for their support in all of my decisions. Thank you so much for instilling the qualities of resilience, kindness, and perseverance in me from a very young age and for being my superheroes.

From the bottom of my heart, I would like to say a big thank you to my beloved sister Dr. Ipshita Upadhyay for her endless provision in completing this and everything else in my life. She is the personification of perfection to me and without her the completion of this thesis or anything else would not have been possible.

A debt of gratitude is also owed to the nicest, most helpful brother-in-law, Mr. Shashank Bajpai. He has been a constant source inspiration whose perpetual encouragement has been priceless.

A special thanks to my friends Isha Singh, Shruti Shaunik, Arushi Nautiyal, and Vaishnavi Dasaka for constantly helping me through challenging times.

Lastly, I would like to thank the Almighty for giving me wonderful parents, teachers, friends, and the perfect milieu. I feel immensely blessed.

Chapter 1 : Introduction

In this thesis, we examine the training and evaluation of board game agents. Our goal is to understand how an existing algorithm, Q-learning, performs as a game playing agent. We provide a methodology to train and study this algorithm for three different abstract board games that increase in complexity, and we provide a detailed methodology on how to train and evaluate Q-learning agents.

Reinforcement Learning (RL) is a powerful model to create learning agents for complex problems. It is a research sub-area in the broader field of artificial intelligence (AI) where, in RL, an agent attempts to maximize the total reward for its actions in an uncertain environment. The application of RL in board games is interesting, because board games present simplified spaces where we can test and observe a decision based agent. Also board games provide a competitive space to compare different AI techniques all within a limited state-space complexity [1]. With a deeper understanding of the performance of RL game agents with respect to changing state-space complexity, we are able to evaluate AI techniques and provide insight on how to employ these agents in a board game market. This includes providing a range of quality AI agents that are challenging for players at all levels of skill, which allows a human player to improve, beat, and have fun playing against.

Our goal is to explore the quantitative difference in the performance of RL based game playing agents with changing state-space complexity of abstract board games. For this purpose, our game playing agents are designed using Q-learning framework for three different games with different state-space complexities. Each agent is trained against a range of AI agents (Min-Max, Q-learning, and Random) to get a better understanding of how quickly we find convergence when training Q-agent for a particular game.

Figure 1.1 shows the details of this thesis. First, we create a Q-learning agent that learns to play three different games - Tic-Tac-Toe, Nine-Men's Morris, and Mancala. Each of these games is of an increased state-space complexity going from first to last. Chapter 2 talks about the different kinds of agents used throughout this thesis, which are deterministic Min-Max agent, non-deterministic Min-Max agent, Q-Learning based agent, and random agent, shown by boxes labelled in yellow in Figure 1.1. The yellow outlined boxes in Figure 1.1 contain the names used to refer each of the agent in the blue outlined boxes that talk about the methodology used for evaluation of the agents, which will be discussed ahead. In chapter 3, boxes labelled in green in Figure 1.1, we analyze our training approach. We report the number of training generations needed when training is done against itself and opponent agents as trained by us. We hypothesize that Q-learning will only be useful as an RL-based technique for a limited state-space complexity, but each of these games chosen has been shown to be solvable, and we do not study this problem beyond our game choices. However, training time increases as state complexity increases. We, also, explore how quickly our agents train to convergence under different teaching agents. From our trials we find that, deterministic Min-Max agent is the best teaching agent in terms of number of training generations required for Tic-Tac-Toe, Nine Men's Morris, and Mancala.

In chapter 3, boxes labelled in green in Figure 1.1, we also show our method to create weakened

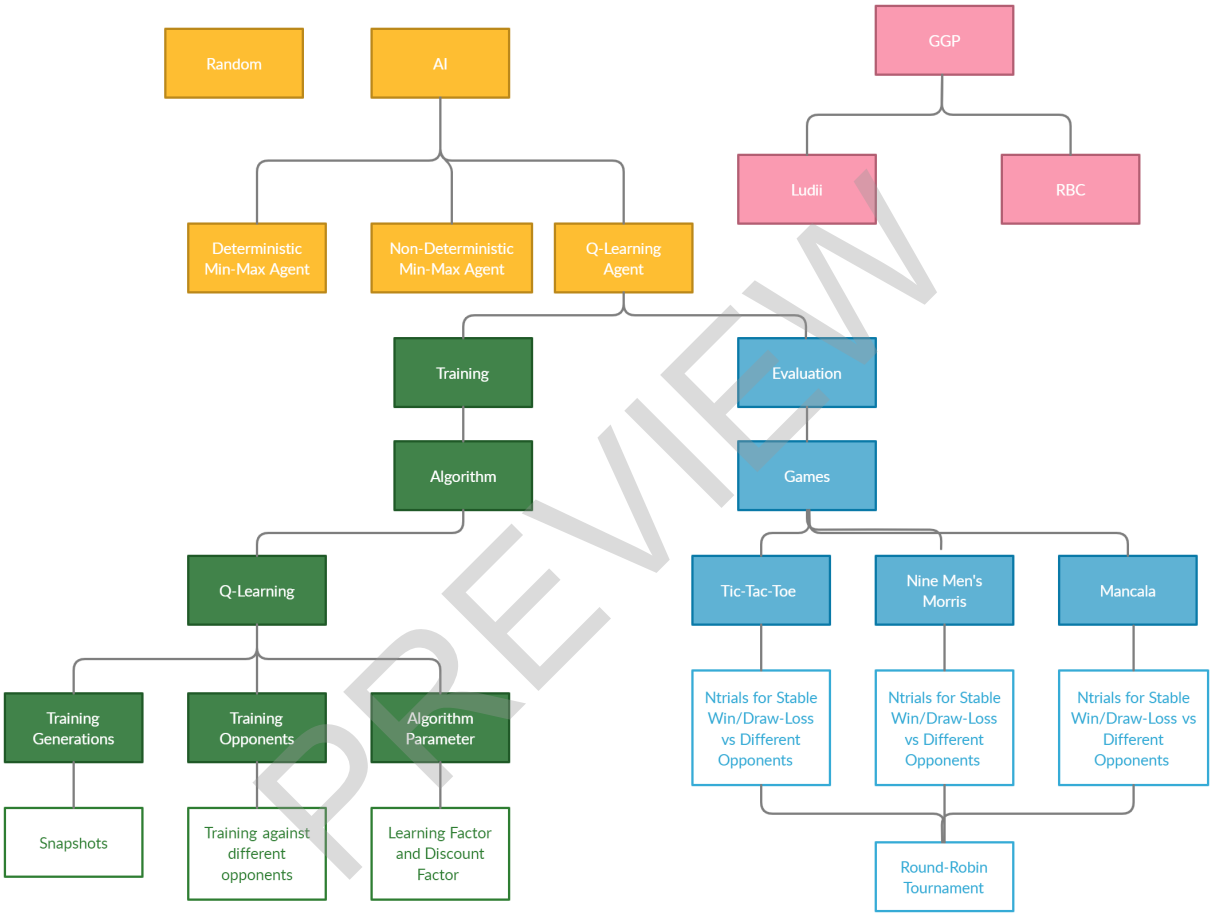

Figure 1.1: The mind map of the research.

agents that will be used in the following chapter. Additionally, to confirm our hypothesis that better teaching agents accelerate training we experiment with training our agent against weakened versions of our Q-learning agent where each teaching agent has increasing number of training generations going from first to last.

Boxes labelled in blue in Figure 1.1 show sections of chapter 4 where we describe our methodology to we compare agents when playing against each other for each of the three games- Tic-Tac-Toe, Nine-Men's Morris, and Mancala. Our approach is a round-robin tournament approach where we can compare agents. To do this, we first need to determine what a reasonable numbers of trial games is needed to conclusively determine the win to draw to loss ratio. We provide our methodology for finding this number of trials for each game, and then use this methodology to compare the range of agents we have against one another.

For this research, we use a General Game Playing (GGP) system, boxes labelled in pink in Figure 1.1, that provide a program that can perform well across various types of games. We specifically use the Ludii General Game System, as created by the Digital Ludeme Project (DLP), that allows us to conduct experiments to investigate if there is a discernible difference in the performance of game playing agents with different state-space complexity. We describe this game playing system in the chapter 2.

The major contributions of this thesis are:

- An analysis of training a Q-learning RL agent in terms of finding convergence for solved zero-sum abstract board games and evaluating training speed based on using a mirror of itself versus a Min-Max agent and a random agent (Chapter 3). This includes how to select parameters in the Q-learning algorithm.

- A method to create lower quality Q-learning based agents to allow human players to find a competitive agent that can provide humans challenging opponents that are not too hard to beat (Chapter 3)

- A method to create a round-robin tournament to evaluate the quality of agents, which includes a method to determine how many games must be played between opponents to find stable results (Chapter 4)

The results of these contributions provide other RL researchers methods to properly evaluate RL agents and use Q-learning based agents in the future for virtual game playing agents.

The thesis is organized as follows: Chapter 2 provides background on the use of RL in gaming, the Ludii General game system, game complexity, the games explored for the purpose of this research, followed by RL approaches for game playing agents including Q-Learning and Min-Max, along with details of the API for developing game-playing agents in Ludii. Chapter 3 describes our training approach and analysis of training Q-learning along with finding the number of generations required by an agent to train for before the algorithm converges on "the best solution", what is the "best solution", and how to decide if an agent is good enough? Additionally, we show how to use the converged agent to create lesser quality agents. Chapter 4 explains how we create a system to evaluate agents. This includes a methodology on how to find the number of trials such that evaluation of agents is stable. We then perform a round-robin tournament for each of the agents