Reinforcement Learing on Board Games

A Project Report Submitted
in Partial Fulfillment of the Requirements
for the Degree of

Bachelor of Technology

by

Vishal Kumar Chaudhary
(111501030)

under the guidance of

Dr. Chandra Shekar Laskhminarayanan

INDIAN INSTITUTE
OF TECHNOLOGY
PALAKKAD

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

*This is to certify that the work contained in this thesis entitled "**Reinforcement Learing on Board Games**" is a bonafide work of **Vishal Kumar Chaudhary** (**Roll No. 111501030**), carried out in the Department of Computer Science and Engineering, Indian Institute of Technology Palakkad under my supervision and that it has not been submitted elsewhere for a degree.*

Dr. Chandra Shekar Laskhminarayanan

Assistant/Associate Professor

Department of Computer Science & Engineering

Indian Institute of Technology Palakkad

Acknowledgements

I would like to express my special thanks to Dr. Chandra Shekar Laskhminarayanan for his support and guidance throughout the course of the project. He gave me various opportunities to learn various things which couldn't be possible all by myself. He gave me ways to organize things and what is the scope of the project. I would also like to thank my family for their support and motivation.

Abstract

In this project, learning through self-play algorithm has been explored and has been applied on connect4, 2048, cricket. Our task is to improve the training process of board games. During self-play, data is being generated for the model to learn. So if we create good data and explore more rewarding paths then our agent can learn quickly. For experiment purposes, we have tried the self-play algorithm in 2048 with a branching factor of 4. Further, in this project we worked on two important things.

We have introduced KL-Upper Confidence Bound (KL-UCB) and Thompson sampling for sequential game connect4. In Spite of these bound came from the solution of multi-arm bandit problem they show significant improvement over the existing bound which was being used in the algorithm.

Learning through self-play algorithm has been implemented for sequential games but we have applied this algorithm on cricket in the simultaneous environment. Simultaneous games are more complex game than its sequential form because there is always randomness involved about the opponent.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Board games are one the best way to test agents and reinforcement learing algorithm to test it viability. Example- chess is one of the complex game with state space of almost 10^{47} and decision tree of size 10^{123}. Similarly state space complexity of connect4 is 10^{12} and decision tree of 10^{21}. The task of this project is to improve the learning rate of the agents who are going to play these games. In this project various games have been chosen depending upon the computation requirement we have.

1.1 Section name

1st Section

1.2 2nd Section name

2nd Section

1.3 Organization of The Report

You can write the about organization of your report in the following manner.

This chapter provides a background for the topics covered in this report. We provided

a description of wireless ad hoc networks, and their applications. Then we described the network model that represents the topology of wireless ad hoc networks [1]. In this chapter it is shown that the virtual backbone for wireless ad hoc networks can be represented by a connected dominating set. We explained clustering concepts and lastly the difference between centralized and distributed algorithms are also discussed. The rest of the chapters are organised as follows: next chapter we provide review of prior works. In Chapter 3 and 4, we discuss our new algorithms for constructing small backbones for ad-hoc wireless network. And finally in chapter 6, we conclude with some future works.

Chapter 2

Review of Prior Works

Survey comes hear

2.1 Section name

write ....

2.2 Conclusion

This chapter provided details of the some of the existing distributed algorithms for constructing a CDS in wireless ad-hoc networks. The results of these evaluations are summarized in table **??**. In next chapter, we discuss our distributed Algorithm I, for constructing a small backbone in ad-hoc wireless network.

Chapter 3

Algorithm I

Game which have been used to test the validity of the algorithm is connect4. Connect4 is a two player game. So the agent have to chose one valid action depending on the board state in such way that he will win the game. This problem is similar to multi-arm bandit problem in which one has to chose arm to maximise the reward or which minimize the regret. During the self play and doing Monte Carlo Tree Search(MCTS) we are using KL-UCB to select action given the history of rewards. In multi-arm bandit problem, KL-UCB has less regret than the UCB. We want to see whether this also work and improves the learning time of agent.

For each state and action of game we store reward during self play and the create data for agent to learn upon those examples.

Step 1: For each state having k valid actions

Step 2: Choose every action once.

Step 3: Then choose A_t action at t - time

$$A_t = argmax_i \ \max \left\{ x \in [0,1] : d(\hat{x}_i(t-1), x) <= \frac{\log N}{N_i} \right\}$$

3.1 Conclusion

In this chapter, we proposed a distributed algorithm for construction of xyz. The complexity of this algorithm is $O(n \log n)$. Next chapter presents another distributed algorithm which has linear time complexity based on xyz.

Chapter 4

Algorithm II

The algorithm presented in previous chapter has $O(n)$ time complexity. We further propose another distributed algorithm in this chapter based on xyz which has linear time complexity.

4.1 Construction

Write ...

4.2 Improved Method

Write...

4.3 Conclusion

In this chapter, we proposed another distributed algorithm for XYZ. This algorithm has both time complexity of $O(n)$ where n is the total number of nodes. In next chapter, we conclude and discuss some of the future aspects.

Chapter 5

Conclusion and Future Work

write results of your thesis and future work.

References

[1] H. A. Omar, K. Abboud, N. Cheng, K. R. Malekshan, A. T. Gamage, and W. Zhuang, "A survey on high efficiency wireless local area networks: Next generation wifi," *IEEE Communications Surveys Tutorials*, vol. 18, no. 4, pp. 2315–2344, Fourthquarter 2016.