

Modeling navigation of Agents using Reinforcement Learning

Laksri Wijerathna

129111G



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Faculty of Information Technology

University of Moratuwa

November 2014

Modeling navigation of Agents using Reinforcement Learning



Laksri Wijerathna
University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Dissertation submitted to the Faculty of Information Technology, University of Moratuwa, Sri Lanka for the partial fulfillment of the requirements of the Degree of M.Sc in Artificial Intelligence

November 2014

Declaration

I declare that this dissertation does not incorporate, without acknowledgment, any material previously submitted for a Degree or a Diploma in any University and to the best of my knowledge and belief, it does not contain any material previously published or written by another person or myself except where due reference is made in the text. I also hereby give consent for my dissertation, if accepted, to be made available for photocopying and for interlibrary loans, and for the title and summary to be made available to outside organization.

Laksri Wijerathna

Name of Student

Signature of Student

Date: _____



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Supervised by

Prof. Asoka Karunananda

Name of Supervisor(s)

Signature of Supervisor(s)

Date: _____

Dedication

This thesis is dedicated to my Father Birnard Michael Wijerathna and my mother the late Muditha Kumari Wijerathna whom I miss every single day. No words are sufficient to describe my father and late mother's contribution to my life. I owe every bit of my existence to her and my father.



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Acknowledgement

A research project like this is never the work of anyone alone. The contributions of many different people, in their different ways, have made this possible. I would like to extend my appreciation especially to the following.

Prof. Asoka Karunanada, for making this research possible. His support, guidance, advice throughout the research project is greatly appreciated. Indeed, without his guidance, I would not be able to put the topic together.

My Dad, Parents in law and Friends, for their unconditional support, both financially and emotionally throughout the journey. In particular, the patience and understanding shown by my dad family and husband during the Study period is greatly appreciated. I know, at times, my temper is particularly trying. Specially, I would like to thank my husband who was always there cheering me up and stood by me through the good times and bad.



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.ho.mrt.ac.lk

Abstract

This thesis addresses the issue of modeling the agent navigation in a benign environment by using reinforcement learning. With the use of reinforcement learning the learning agent is created who does not have any idea about the environment that the agent is placed initially. The agent will be exploring the environment and with the use of the rewards and penalties the agent perceives from the environment the agent will be learning how to act to maximize the rewards that agent is getting from the environment. To implement the learning process Temporal Difference learning algorithm Q-learning and Sarsa is used. The system uses the user defined source and goal as the input to the system and output the optimal navigation path from the given source to the goal by learning the environment itself. The learning agent is update their knowledge and will find a policy value for the reward function and the value function. The main objective of this thesis is to give an approach in which the agent learning happens without any human intervention The most important part of this learning agent is at the time of initiation agent is not having any idea about the environment but will learn eventually on how to navigate in a benign environment without having any initial knowledge of the environment. The output of the system is the optimal path generated by the learning agent form the given source to the goal. To evaluate the system the results of the agent behavior is checked against the human knowledge mapping on navigation.

Table of contents

Chapter 1 Introduction	1
1.1 Prolegomena	1
1.2 Outline	4
Chapter 2 Related work in Agents and Reinforcement Learning	6
2.1 Introduction	6
2.2 Major research in Agents and Machine Learning	6
2.2.1 Grounded Language Learning	8
2.2.2 Reinforcement learning with multi agent systems	11
2.3 Major research in Neural Network and Reinforcement Learning	14
2.4 Summary	16
Chapter 3 Reinforcement learning for agent navigation	17
3.1 Introduction	17
3.2 Reinforcement learning	18
3.2.1 Elements of reinforcement learning	19
3.2.2 Temporal Difference Learning	20
3.3 Intelligent Agents and their behavior	22
3.4 Summary	25
Chapter 4 Use of Reinforcement learning for agent navigation	26
4.1 Introduction	26
4.2 Hypothesis	27
4.3 Input	28
4.4 Output	28
4.5 Process	28



4.6 Features	33
4.7 Evaluation	33
4.8 Summary	33
Chapter 5 Design	34
5.1 Introduction	34
5.2 System Design	34
5.3 Summary	38
Chapter 6 Implementation	39
6.1 Introduction	39
6.2 System Implementation	39
6.3 Summary	48
Chapter 7 Evaluation	49
7.1 Introduction	49
7.2 Evaluation Process	49
7.2 Conclusion	55
Chapter 8 Conclusion	56
8.1 Future work	57
References	58
Appendix A: Intelligence Agent and their behaviour	62
A.1 Intelligence Agent and their behaviour comparison	62
A.2 Intelligence Agents and the structure of the agents	63
A.3 Ant-Q algorithm	69
A.4 Deep fitted Q algorithm schema	70
Appendix B: Major Implementation in Software Development	71
B.1 Implementation of the RLearner Class	71



University of Moratuwa, Sri Lanka.
 Electronic Theses & Dissertations
www.lib.mrt.ac.lk



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

List of Figures

Figure 3.1: Value function	20
Figure 3.2: Q-value iteration for Sarsa algorithm	21
Figure 3.3: Q-value iteration for Q-learning algorithm	22
Figure 4.1: A visual field of the agent	29
Figure 4.2: Agent navigation environment	30
Figure 4.3: A visual field of the goal to the agent	31
Table 4.1: Visual field and perceptual state representation on the agent	32
Figure 5.1: Agents basic interaction with the system	34
Figure 5.2: System Architecture diagram	36
Figure 5.3: Offline training diagram of the agent	38
Figure 6.1: Class Diagram	40
Figure 6.2: RLWorld interface	41
Figure 6.3: Q-learning algorithm	42
Figure 6.4: Q-learning algorithm implementation	43
Figure 6.5: Sarsa algorithm	44
Figure 6.6: Sarsa algorithm implementation	45
Figure 6.7: ϵ -greedy policy action selector implementation	47
Figure 6.8: softmax policy action selector implementation	48
Figure 7.1: Visualization of Agents accuracy over number of epochs for Sarsa algorithm	52
Figure 7.2: Visualization of average accuracy over number of epochs for Sarsa algorithm	53
Figure 7.3: Visualization of Agents accuracy over number of epochs for Q-learning algorithm	54
Figure 7.4: Visualization of average accuracy over number of epochs for Q-learning algorithm	54
Figure 7.5: Visualization of average accuracy for Q-learning algorithm vs. Sarsa algorithm	55

List of Tables

Table 4.1: Visual field and perceptual state representation on the agent	32
Table 7.1: Agent Accuracy when using the Q-learning algorithm for learning	50
Table 7.2: Agent Accuracy when using the Q-learning algorithm for learning	51



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk