

**A HYBRID APPROACH FOR DYNAMIC TASK
SCHEDULING IN UNFORESEEN ENVIRONMENTS USING
MULTI AGENT REINFORCEMENT LEARNING AND
ENHANCED Q-LEARNING**

Jayakody Kankanamalage Chathurangi Shayalika
189394E

Degree of Master of Science in Artificial Intelligence

Department of Computational Mathematics

University of Moratuwa
Sri Lanka

October 2020

**A HYBRID APPROACH FOR DYNAMIC TASK
SCHEDULING IN UNFORESEEN ENVIRONMENTS USING
MULTI AGENT REINFORCEMENT LEARNING AND
ENHANCED Q-LEARNING**

Jayakody Kankanamalage Chathurangi Shayalika
189394E

Thesis submitted in partial fulfilment of the requirements for the
degree of Master of Science in Artificial Intelligence

Department of Computational Mathematics

University of Moratuwa
Sri Lanka

October 2020

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.

Name of Student

J.K.C. Shayalika

Signature of Student

Date: 30/10/2020

Supervised by

Name of Supervisor(s)

Dr. Thushari Silva

Signature of Supervisor(s)

Date: 30/10/2020

Dedication

I dedicate this thesis to my parents who are the pioneers and great pillars that reinforced my education and always supported and inspired me. This thesis is gratefully dedicated to all the lecturers of Department of Computational Mathematics, Faculty of Information Technology, University of Moratuwa.

Acknowledgement

I would like to express my sincere gratitude to my supervisor Dr. Thushari Silva who guided and profoundly nurtured me throughout the project. She extended her fullest corporation to me whenever I sought for advice. I'm also thankful to Prof. Asoka Karunananda who helped immensely from the beginning of the research. I was encouraged by his teaching methods to consider issues in multiple avenues.

I'm grateful to all the lecturers of Department of Computational Mathematics. During the course period of two years, these valuable lectures helped me to think differently.

I had to refer to several books and research papers as references for this research. I want to thank all the authors of these publications for their notable work. I'm especially thankful to "*Commonwealth of Learning-Canada*", online professional development program who assisted the fully access to online Specialization on Reinforcement Learning offered by *University of Alberta, Canada*, where I could grasp unparalleled knowledge on Reinforcement Learning.

Many thanks to all the editors and reviewers of the journals and conferences whom I submitted the papers related to the project, whose positive and negative feedback greatly assisted the progress of the research.

And, in many ways, my batch mates helped me a lot. I would like to give them my respect.

I want to express my deepest gratitude to my caring parents and sister for always helping me to pursue higher education.

Abstract

The process of assigning most appropriate resources to workstations or agents at the right time is termed as Scheduling. The word is applied separately to tasks and resources in task scheduling and resource allocation accordingly. Scheduling is a universal theme being conferred in technological areas like computing and strategic areas like operational management. The core idea behind scheduling is the distribution of shared resources across time for competitive tasks. Optimization, efficiency, productivity and performance are the major metrics evaluated in scheduling. Effective scheduling under uncertainty is tricky and unpredictable and it's an interesting area to study. Environmental uncertainty is a challenging extent that effect scheduling based decision making in work environments where environment dynamics subject to numerous fluctuations frequently.

Reinforcement Learning is an emerging field extensively research on environmental modelling under uncertainty. Optimization in dynamic scheduling can be effectively handled using Reinforcement learning. This research is about a research study that focused on Reinforcement Learning techniques that have been used for dynamic task scheduling. This thesis addresses the results of the study by means of the state-of-the-art on Reinforcement learning techniques used in dynamic task scheduling and a comparative review of those techniques. This thesis reports on our research on a Hybrid Approach for Dynamic Task Scheduling in Unforeseen Environments using the techniques; Multi Agent Reinforcement Learning and Enhanced Q-Learning.

The proposed solution follows online and offline reinforcement learning approaches which works on real time inputs of heuristics like, Number of agents involved, current state of the environment and backlog of tasks and sub-tasks, Rewarding criteria etc. The outputs are the set of scheduled tasks for the work environment. The solution comes with an approach for priority based dynamic task scheduling using Multi Agent Reinforcement Learning & Enhanced Q-Learning. Enhanced Q-Learning includes developed algorithm approaches; Q-Learning, Dyna Q+ Learning and Deep Dyna-Q+ Learning which is proposed as an effective methodology for scheduling problem.

The novelty of the solutions resides on implementation of model-based reinforcement learning and integration with the model-free reinforcement learning algorithmic approach by means of Dyna-Q+ Learning and Deep Dyna-Q+ Learning for dynamic task scheduling in an unforeseen environment. The research project also concentrates on how the dynamic task scheduling is managed within a constantly updating environment which the Deep Dyna-Q+ has provided a ground solution to cater this requirement. The end solution has comparatively evaluated the product using evaluation metrics in each of the three Q-Learning variations developed. As per the evaluation results it was revealed Deep Dyna-Q+ implementation would cater well the problem of dynamic task scheduling in an unforeseen environment.

Table of Contents

Declaration	iii
Dedication	iv
Acknowledgement	v
Abstract	vi
Table of Contents	vii
List of Figures	xii
List of Tables	xiv
Abbreviations	xv
Chapter 01 Introduction	1
1.1 Prolegomena.....	1
1.2 Background and Motivation.....	2
1.3 Problem in Brief.....	3
1.4 Aim.....	4
1.5 Objectives.....	4
1.6 Hypothesis.....	4
1.7 Solution	4
1.8 Structure of Thesis	4
1.9 Summary	5
Chapter 02 Evolution and State-of-the-art of Reinforcement Learning in Dynamic Task Scheduling.....	6
2.1 Introduction	6
2.2 Breakthroughs in Reinforcement scheduling	6
2.3 Model-Free RL.....	7
2.3.1 Value-Iteration Methods.	7
2.3.2 Policy Optimization or Policy-Iteration Methods.....	11
2.4 Model-Based RL	12
2.5 Integration of Model-Free and Model-Based RL.....	13
2.5.1 Dyna-Q Algorithm	13
2.6 Standard Dynamic Task/Resource Allocation Frameworks	13
2.6.1 MARL (Multi-Agent Reinforcement Learning).....	13
2.6.2 OSL (Ordinal Sharing Learning).....	16
2.6.3 Gossip-Based Reinforcement Learning (GRL).....	17
2.6.4 Centralized Learning Distributed Scheduling (CLDS)	17

2.7 Research Findings	18
2.8 Challenges in Reinforcement scheduling	20
2.9 Problem Definition	20
2.10 Summary	21
Chapter 03 Technology.....	22
3.1 Introduction	22
3.2 Reinforcement Learning.....	22
3.3 Elements of Reinforcement Learning	22
3.3.1 Agents.....	22
3.3.2 Actions.....	23
3.3.3 States.....	23
3.3.4 Environment	23
3.3.5 Policy	23
3.3.6 Reward Signal.....	23
3.3.7 Value Function	23
3.3.8 Environment Model.....	23
3.4 Markov Decision Processes	24
3.5 Model-Free RL & Model-Based RL	24
3.6 Model-Free RL.....	24
3.6.1 Q-Learning.....	24
3.6.2 Deep Q-Learning	26
3.7 Model-Based RL	27
3.8 Integration of Model-free RL & Model-based RL.....	27
3.8.1 Dyna –Q Algorithm.....	27
3.8.2 Dyna – Q+ Algorithm.....	28
3.9 Action Selection Mechanisms.....	29
3.10 Multi Agent Systems.....	30
3.11 Single Agent Reinforcement Learning and Multi Agent Reinforcement Learning	31
3.12 Types of Agents in a MARL System	32
3.14 Summary	32
Chapter 04 Approach	33
4.1 Introduction	33
4.2 Hypothesis.....	33
4.3 Input	33
4.4 Output.....	33

4.5	Process.....	33
4.6	Features	35
4.7	Users.....	35
4.8	Summary	35
Chapter 05 Design.....		36
5.1	Introduction	36
5.2	System Design.....	36
5.3	System Model.....	36
5.3.1	Agents	36
5.3.2	Product Backlog and Tasks.....	36
5.3.3	Rewards.....	37
5.3.4	Time	37
5.3.5	Environment.....	37
5.3.6	Actions	37
5.3.7	State.....	37
5.4	System Architecture	38
5.5	Modular Architecture	38
5.5.1	Data Acquisition Module.....	39
5.5.2	MARL Module	39
5.5.3	Scheduler Module.....	40
5.5.4	Environment Module.....	41
5.5.5	Result Logging Module	41
5.6	Simulation Design.....	41
5.7	Summary	42
Chapter 06 Implementation		43
6.1	Introduction	43
6.2	Implementation of Modules	43
6.2.1	Data Acquisition Module.....	43
6.2.2	Environment Module	44
6.2.3	Scheduler Module.....	45
6.2.4	MARL Module	49
6.2.5	Result Logging Module	50
6.3	Action Selection Mechanisms Developed.....	50
6.4	Generalization to Unseen Environments.....	51
6.4.1	System Level.....	51

6.4.2	User Level.....	51
6.5	Tools and Technologies	52
6.5.1	Integrated Development Environments (IDEs) Used.....	52
6.5.2	Languages Used.....	52
6.5.3	Libraries Used.....	52
6.5.4	Hardware Used	53
6.6	Summary	53
Chapter 07	Evaluation.....	54
7.1	Introduction	54
7.2	Experimental Design	54
7.2.1	Dyna Q+ Learning Simulation Results.....	55
7.2.2	Deep Dyna-Q+ Network Learning Simulation Results.....	56
7.2.3	Visualization of Task Schedule	56
7.3	Analysis of Results.....	58
7.3.1	Reward Analysis	58
7.3.2	Evaluation of the Throughput in Deep Dyna-Q+ learning.	59
7.3.3	Evaluation of Variation of output with number of agents in Deep Dyna-Q+ learning	60
7.3.4	Evaluation of Variation of output with performance between Random learning, Reinforcement learning, DQ learning and Dyna-Q+ learning	61
7.3.5	Cost Analysis	61
7.4	Summary	62
Chapter 08	Conclusion & Further Work	63
8.1	Introduction	63
8.2	Conclusion.....	63
8.3	Project Results Overview	64
8.4	Limitations and Further Work.....	64
8.5	Summary	65
References	66
Appendix 01:	Sample Codes	69
Appendix 02:	Class Diagram	83
Appendix 02.A:	Class Diagram for Dyna Q+ Learning.....	83
Appendix 02.B:	Class Diagram for Deep Dyna Q+ Learning.....	84
Appendix 03:	Publications	85
Appendix 03.A:	First Page of the Review Paper.....	85

Appendix 03.B: First Page of the Research Paper.....86

List of Figures

Figure 2.1 : Classification of Reinforcement Learning Techniques used in Dynamic Task Scheduling.	7
Figure 2.2 : Deep Q-Learning Driven Framework for Task Allocation. [18]	10
Figure 2.3 : Illustration of multi-UAV communication networks [31].	14
Figure 2.4 : The Architecture of SchedNet [33]	16
Figure 2.5 : The Schematic Diagram of the OSL Method for Job Scheduling [35].....	16
Figure 3.1: Main Algorithmic Difference between Q learning vs Deep Q learning	26
Figure 3.2: Relationship Among Learning, Planning and Acting [14]	28
Figure 3.3: The General Dyna Architecture [14]	28
Figure 3.4: Exploration and Exploitation behavior in e-epsilon-decay	30
Figure 3.5: Single Agent RL [14]	31
Figure 3.6: Multi Agent RL [14]	31
Figure 4.1: Proposed System Flowchart.....	34
Figure 5.1: System Architecture.	38
Figure 5.2: Modular Architecture	39
Figure 5.3: Simulation Design.....	42
Figure 6.2: Sample Environment Generated (for 3 tasks, 2 agents).....	44
Figure 6.3: Sample Task Backlog.....	45
Figure 6.4: Reward Function Implemented	45
Figure 6.5: RL Q-Learning Process.....	46
Figure 6.7: DQN Function Approximation	48
Figure 6.8: Structure of DQN Developed.....	49
Figure 6.9: Final Scheduled environment.....	49
Figure 6.10: Agent Communication, Tasks Co-ordination	50
Figure 7.1: Representation of Job Completions	55
Figure 7.3: Agent Interaction, States and Actions.....	55
Figure 7.2: Representation of Job Completions	55
Figure 7.4: Saved Q-Table.....	55
Figure 7.5: Agent Message Passing, States Representation and Job Completions in Deep Dyna-Q+	56
Figure 7.6: Sample Task Schedule Resulted from Deep Dyna-Q+ Simulation	57

Figure 7.7: Isolating Traces (such that flow of only needed tasks can be viewed)	57
Figure 7.8: Lasso Select and Compare Data on Hover.....	58
Figure 7.9: Plot of Reward Analysis I	59
Figure 7.11: Cost Analysis of an instance in Dyna-Q+ Simulation	62
Figure 7.12: Cost Analysis of an instance in Deep Dyna-Q+ Simulation.....	62

List of Tables

Table 2.1 : Comparison of Reinforcement Learning Techniques used for Dynamic Task Scheduling.	18
Table 2.2: Comparison of Reinforcement Learning Based Task Scheduling Frameworks	19
Table 7.1: Task Backlog -Evaluation I.....	59
Table 7.2: Results-Evaluation II.....	59
Table 7.3: Task Backlog -Evaluation III	60
Table 7.4: Results-Evaluation III.....	60
Table 7.5: Task Backlog -Evaluation IV	61
Table 7.6: Results- Evaluation IV	61
Table 8.1: Project Results Overview	64

Abbreviations

DL	-	Deep Learning
RL	-	Reinforcement Learning
DRL	-	Deep Reinforcement Learning
QL	-	Q Learning
DQL	-	Deep Q Learning
DQN	-	Deep Q-Network
MDP	-	Markov Decision Process
MARL	-	Multi Agent Reinforcement Learning
DDQ+	-	Deep Dyna Q+ Learning
GPU	-	Graphics Processing Unit