University of Newcastle

Doctoral Thesis

Cooperative Reinforcement Learning for Independent Learners

Author:
Bilal Hashem Kalil Abed-Alguni
BCS, MCS

A thesis submitted to the
University of Newcastle, NSW, Australia
for the degree of
Doctor of Philosophy
(Computer Science)

October 2014

Statement of Originality

The thesis contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. I give consent to the final version of my thesis being made available worldwide when deposited in the University's Digital Repository**, subject to the provisions of the Copyright Act 1968. **Unless an Embargo has been approved for a determined period.

(Bilal Hashem Abed-Alguni)

Contents

List of Figures

Abstract

Cooperative Reinforcement Learning for Independent Learners

by Bilal Abed-Alguni, BCS, MCS

University of Newcastle
Faculty of Engineering and Built Environment
School of Electrical Engineering and Computer Science

A thesis submitted for the degree of Doctor of Philosophy

Machine learning in multi-agent domains poses several research challenges. One challenge is how to model cooperation between reinforcement learners. Cooperation between independent reinforcement learners is known to accelerate convergence to optimal solutions. In large state space problems, independent reinforcement learners normally cooperate to accelerate the learning process using decomposition techniques or knowledge sharing strategies. This thesis presents two techniques to multi-agent reinforcement learning and a comparison study. The first technique is a formal decomposition model and an algorithm for distributed systems. The second technique is a cooperative Q-learning algorithm for multi-goal decomposable systems. The comparison study compares the performance of some of the best known cooperative Q-learning algorithms for independent learners.

Distributed systems are normally organised into two levels: system and subsystem levels. This thesis presents a formal solution for decomposition of Markov Decision Processes (MDPs) in distributed systems that takes advantage of the organisation of distributed systems and provides support for migration of learners. This is accomplished by two proposals: a Distributed, Hierarchical Learning Model (DHLM) and an Intelligent Distributed Q-Learning algorithm (IDQL) that are based on three specialisations of agents: workers, tutors and consultants. Worker agents are the actual learners and performers of tasks, while tutor agents and consultant agents are coordinators at the subsystem level and the system level, respectively. A main duty of consultant and tutor agents is the assignment of problem space to worker agents. The experimental results in a distributed hunter prey problem suggest that IDQL converges to a solution faster than the single agent Q-learning algorithm. An important feature of DHLM is that it provides a solution for migration of agents.

This feature provides support for the IDQL algorithm where the problem space of each worker agent can change dynamically. Other hierarchical RL models do not cover this issue.

Problems that have multiple goal-states can be decomposed into sub-problems by taking advantage of the loosely-coupled bonds among the goal states. In such problems, each goal state and its problem space form a sub-problem. This thesis introduces Q-learning with Aggregation algorithm (QA-learning), an algorithm for problems with multiple goal-states that is based on two roles: learner and tutor. A learner is an agent that learns and uses the knowledge of its neighbours (tutors) to construct its Q-table. A tutor is a learner that is ready to share its Q-table with its neighbours (learners). These roles are based on the concept of learners reusing tutors' sub-solutions. This algorithm provides solutions to problems with multiple goal-states. In this algorithm, each learner incorporates its tutors' knowledge into its own Q-table calculations. A comprehensive solution can then be obtained by combining these partial solutions. The experimental results in an instance of the shortest path problem suggest that the output of QA-learning is comparable to the output of a single Q-learner whose problem space is the whole system. But the QA-learning algorithm converges to a solution faster than a single learner approach.

Cooperative Q-learning algorithms for independent learners accelerate the learning process of individual learners. In this type of Q-learning, independent learners share and update their Q-values by following a sharing strategy after some episodes learning independently. This thesis presents a comparison study of the performance of some famous cooperative Q-learning algorithms (BEST-Q, AVE-Q, PSO-Q, and WSS) as well as an algorithm that aggregates their results. These algorithms are compared in two cases: equal experience and different experiences cases. In the first case, the learners have equal learning time, while in the second case, the learners have different learning times. The comparison study also examines the effects of the frequency of Q-value sharing on the learning speed of independent learners.

The experimental results in the equal experience case indicate that sharing of Q-values is not beneficial and produces similar results to single agent Q-learning. While, the experimental results in the different experiences case suggest that each of the cooperative Q-learning algorithms performs similarly, but better than single agent Q-learning. In both cases, high-frequency sharing of Q-values accelerates the convergence to solutions compared to low-frequency sharing. Low-frequency Q-value

sharing degrades the performance of the cooperative Q-learning algorithms in the equal experience and different experiences cases.

Acknowledgements

This Ph.D thesis would not have seen the light without the help and support of the kind people around me, to only some of whom it is possible to give particular mention here.

First and foremost I offer my sincerest gratitude to my supervision team: A/Prof. Stephan Chalup, A/Prof. Frans Henskens, A/Prof. Huilin Ye, Dr David Paul, and Dr Yuqing Lin for the continuous support of my Ph.D research. Their patience, advice, encouragement, motivation, and enthusiasm have been of great value to me. Indeed, without their guidance, I would not be able to put the topic together.

I would like to acknowledge the financial, academic and technical support of the University of Newcastle and its staff, particularly in the award of University of Newcastle International Postgraduate Research Scholarship (UNIPRS) and University of Newcastle Research Scholarship Central 5050 that provided the necessary financial support for this research. The library and computer facilities of the university have been indispensable for the completion of my research and writing my doctoral thesis.

I would like to thank everybody who has been involved in the Interdisciplinary Machine Learning Research Group (IMLRG), and the Distributed Computing Research Group (DCRG).

I would like to thank my confirmation committee: A/Prof. Frans Henskens, Dr Yuqing Lin, and Dr Alexandre Mendes. I admit that my research would be much poorer without the constructive criticism and advice that were provided during my confirmation exam.

Finally, and most importantly, I would like to thank my family for their unconditional support, especially my wonderful wife Flordeliza, my lovely daughter Rua, my parents Hashem and my departed mother Sukiba, my sisters Yara Rua Lara Maha, and my brother Tareq. Your encouragement and emotional support helped me to overcome the hard moments that I have been through during my research. I will always owe them a debt of gratitude for believing in me, even when I did not. A smile from baby Rua was enough to encourage me and meant the world to me.

Abbreviations

ABM	Agent-Based Model
AC	Actor-Critic
AMRLS	Aggregated Multiple Reinforcement Learning System
AOEs	Areas Of Expertise
AVE-Q	Average Q-learning
BA	Boltzman Addition
BEST-Q	BEST Q-learning
BM	Boltzman Multiplication
CBR	Case-Based Reasoning
CG	Coordination Graph
CST	Constructing Skill Trees
DFG	Dissolution and Formation of Groups
DHLM	Distributed Hierarchical Learning Model
FMDP	Factored Markov Decision Process
FMQ	Frequent Maximum Q-values
FIRL	Feature Construction for Inverse Reinforcement Learning
MARL	Multi-agent Reinforcement Learning
MAS	Multi-agent System
MCLA	Multi-agent Cooperative Learning algorithm
MCLM	Multi-agent Cooperative Learning Model
MDP	Markov Decision Process
MV	Majority Voting
IDQL	Intelligent Distributed Q-Learning

POMDP	Partially Observable Markov Decision Process
PSO-Q	Particle Swarm Optimisation Q-learning
QA-learning	Q-learning by Aggregation
RL	Reinforcement Learning
RV	Rank Voting
SARSA	State-Action-Reward-State-Action
SCMARL	State-clusters Multi-agent Reinforcement Learning
SRL	Selfish Reinforcement Learning
WSS	Weighted Strategy Sharing
WMV	Weighted Majority Voting
2LRL	Two-Level Reinforcement Learning

Dedicated to my family.