The University of Newcastle, Australia

Doctoral Thesis

Deep Reinforcement Learning for Cognitive Radio and Software-Defined Networks

---

Syed Qaisar Jalil
BSc (Telecommunication and Networking)

A thesis submitted in partial fulfilment of the requirements or the degree of Doctor of Philosophy in Computer Science

in the

Interdisciplinary Machine Learning Research Group

School of Information and Physical Sciences

College of Engineering, Science and Environment

December 2021

In memory of my father

To my mother
with love and eternal appreciation

Statement of Originality

I hereby certify that the work embodied in the thesis, titled 'Deep Reinforcement Learning for Cognitive Radio and Software-Defined Networks', is my own work, conducted under normal supervision. The thesis contains no material which has been accepted, or is being examined, 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. 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 and any approved embargo.

Signed: Syed Qaisar Jalil

Date: 27/02/2021

Acknowledgement of Authorship

I, Syed Jalil, hereby certify that the work embodied in this thesis contains published and submitted papers of which I am a joint author. I declare that I have contributed to the following papers,

- **S.Q. Jalil**, M.H.Rehmani, S. Chalup, "DQR: Deep Q-Routing in Software Defined Networks", in International Joint Conference on Neural Networks (IJCNN), Virtual on July 19-24, 2020. Publisher: IEEE, DOI: IJCNN48605.2020.9206767.
- **S.Q. Jalil**, M.H.Rehmani, S. Chalup, "A Deep Reinforcement Learning Approach to Fair Distributed Dynamic Spectrum Access", in International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MobiQuitous), Virtual on December 7-9, 2020. Publisher: ACM, DOI: 10.1145/3448891.3448935.
- **S.Q. Jalil**, M.H.Rehmani, S. Chalup, "Cognitive Radio Spectrum Sensing and Prediction Using Deep Reinforcement Learning", in International Joint Conference on Neural Networks (IJCNN), Virtual on July 19-24, 2021. Publisher: IEEE, DOI: IJCNN52387.2021.9533497.
- **S.Q. Jalil**, M.H.Rehmani, S. Chalup, "Spectrum Sensing in Cognitive Radio Networks using Offline Deep Reinforcement Learning" (Under review).

by,

- contributing to each study's conception and design,
- developing analysis plans,
- developing research material and collecting data,
- performing both quantitative and qualitative analyses,

- writing code and execution of programs,
- interpreting data,
- and leading the writing of the manuscripts.

Supervisor endorsement:

By signing below I, Dr Stephan Chalup, confirm that my PhD student Syed Qaisar Jalil contributed to the above listed papers as stated,

Signed: ______________________________

Date: 27-02-2021

Acknowledgements

First and foremost, All praise to Allah the Almighty, who gave me strength and endurance to complete this thesis. May Allah's blessing goes to His final Prophet Muhammad (peace be up on him), his family and his companions.

I would like to express my sincere gratitude to my supervisor, A/Prof. Stephan K Chalup, for his time, patience, and generous guidance throughout my PhD, from which I have learned so much.

I am extremely thankful to Mufti Naeem sb, Dr. Mubashir Husain Rehmani, and Dr. Yasir Saleem who not only motivated me to pursue my doctoral degree but also guided me throughout my PhD with their valuable input and encouragement.

I would also like to thank Geoff Martin for providing me with technical support from the latest Alienware computers to HD monitors.

I am also grateful to my cricket team-mates for creating such wonderful memories and providing a way to reduce my stress levels when the going got tough during my thesis. Special thanks to Uzair Khan, Asad Abbas, Abdul Jabbar, Arif Hussain, and Abdul Khan for their relentless support all the way from the cricket pitch to the lab computer.

I cannot leave the University of Newcastle without mentioning Sameer Ahmad, Abdul Mannan, Gulam Dastagir Khan, Muhammad Javed and Prakash Mallick. Thank you for the lively discussions, constructive comments, and practical suggestions.

Many thanks to my friend, Faheem Ahmad Qureshi, for his encouragement throughout the duration of this thesis.

Finally, I would like to express my deepest gratitude to my family for their patience, tolerance, continuous support, and encouragement. To my beloved father, Syed Ayub Hussain Shah Bukhari, who died last year; I miss you and I wish you were here with us. To my mother, Nighat Yasmin, thank you for the untold number of sacrifices you have made for the entire family, and specifically for me to continue my schooling. To my brother, Syed Awais Bukhari, and sister Sara, thank you for your support. This thesis would never have been completed without you all. I thank you all.

اظہارِ تشکر

تمام تعریفیں اللہ ہی کے لئے ہیں جس نے مجھے توفیق اور ہمت دی کہ میں پی ایچ ڈی کا مقالہ مکمل کر سکوں۔ لاکھوں درود سلام ہوں نبی آخر الزمان محمد ﷺ، آپ کے اہل بیت اور صحابہ کرام پے۔

* میں اپنے پروفیسر سٹیفن چالپ کا انتہائی مشکور ہوں جن کی سرپرستی اور رہنمائی کے بغیر پی ایچ ڈی مکمل کرنا نا ممکن تھا۔

* میں مفتی نعیم میمن صاحب کا بھی تہہ دل سے شکر گزار ہوں جنہوں نے مجھے اپنا قیمتی وقت دیا اور ایک اچھا محقق اور اچھا مسلمان بننے میں میری رہنمائی کی۔

* میں ڈاکٹر مبشر حسین رحمانی کا بے حد مشکور ہوں جنہوں نے نہ صرف مجھے ڈاکٹریٹ کی ڈگری حاصل کرنے کی ترغیب دی بلکہ اپنے قیمتی ان پٹ اور حوصلہ افزائی سے پی ایچ ڈی کے دوران میری رہنمائی بھی کی۔ آپ کی رہنمائی کی وجہ سے پی ایچ ڈی کے سفر کی مشکلات کا سامنا کرنے میں مجھے مدد ملی۔

* میرے دوست، ڈاکٹر یاسر سلیم اور فہیم احمد قریشی کا بہت شکریہ، جنہوں نے اس مقالے کی پوری مدت میں حوصلہ افزائی کی۔

* میں جیوف مارٹن کا بھی شکریہ ادا کرنا چاہوں گا کہ انہوں نے مجھے جدید ترین ایلین ویئر کمپیوٹرز سے لے کر ایچ ڈی مانیٹر تک تکنیکی مدد فراہم کی۔

* گھر سے دور ہونے کے بعد اور ایک انجان ملک میں تنہائی کے دور میں میرے وہ ساتھی جن کی وجہ سے مجھے حوصلہ اور ہمت ملی اور احساسِ تنہائی کم ہوا ان کا بھی شکر گزار ہوں جن میں سر فہرست عزیر خان، اسد عباس، عبد الجبار، محمد جاوید، عبد المنان، سمیر احمد، غلام دستگیر اور پرکاش ملک شامل ہیں۔

* آخر میں، میں اپنی فیملی کا شکریہ ادا کرنا چاہوں گا جن کے بغیر میرا پی ایچ ڈی کرنا نا ممکن تھا۔ میرے والد سید ایوب حسین شاہ بخاری، جن کا انتقال 2019 میں ہوا، کی خواہش تھی کہ ہم سب بہن بھائی اعلیٰ تعلیم حاصل کریں۔ آج مجھے فخر ہے کہ اللہ کے کرم سے اور والدین کی دعاؤں سے میں نے پی ایچ ڈی کی ڈگری حاصل کی اور اپنے والد کی خواہش کو پورا کیا۔ کاش کے آپ آج ہم میں ہوتے، آپ کی کمی ہمیشہ ہماری زندگیوں میں رہے گی۔ جہاں میرے والد کی انتھک محنت ہے وہاں ہی میری والدہ (نگہت یاسمین) کی بھی اتنی ہی قربانیاں اور دعائیں ہیں، میں آپ کا بھی اتنا ہی شکر گزار ہوں۔ میں اپنے بھائی (سید اویس بخاری) اور بہن (سارہ) کا بھی شکریہ ادا کرنا چاہوں گا جنہوں نے اس سفر میں میرا ساتھ دیا۔

Contents

List of Figures

List of Tables

List of Abbreviations

ACER Actor Critic with Experience Replay. 135

ANN Artificial Neural Network. 52

BS Base Station. 24

CM Covariance Matrix. 52

CNN Convolutional Neural Network. 57

CQL Conservative Q-Learning. 31

CR Cognitive Radio. 22

CRN Cognitive Radio Network. 23

CSS Cooperative Spectrum Sensing. 28

CTMC Continuous-Time Markov Chain. 102

D3QN-PER Dueling Double Deep Q-Networks with Prioritised Experience Replay. 31

DCQL Discrete Conservative Q-Learning. 43

DCS Deep Cooperative Sensing. 97

DCS-HD Deep Cooperative Sensing Hard. 97

DCS-SD Deep Cooperative Sensing Soft. 97

DDPG Deep Deterministic Policy Gradient. 162

DL Deep Learning. 20

DNN Deep Neural Network. 39

DQN Deep Q-Network. 39

DQR Deep Q-Routing. 32

DRL Deep Reinforcement Learning. 20

DSA Dynamic Spectrum Access. 22

DTMC Discrete-Time Markov Chain. 101

ED Energy Detection. 51

FC Fusion Center. 28

GOOSE Generic Object-Oriented Substation Even. 168

GPD Generalised Pareto Distribution. 102

HMM Hidden Markov Model. 56

IQL Independent Q-learning. 41

LARAC Lagrangian relaxation-based aggregated cost. 168

LSS Local Spectrum Sensing. 27

LSTM Long Short-Term Memory. 107

MDP Markov Decision Processes. 34

ML Machine Learning. 19

NBIs Northbound Interfaces. 145

NOMA Non-Orthogonal Multiple Access. 50

OOD Out-of-distribution. 42

PDF Probability Density Function. 102

PER Prioritised Experience Replay. 41

PPO Proximal Policy Optimisation. 82

PUs Primary Users. 23

QoS Quality-of-service. 21

RL Reinforcement Learning. 19

SBIs Southbound Interfaces. 145

SDN Software-defined Network. 25

SNR Signal-to-noise. 57

SOP Spectrum Occupancy Prediction. 28

SP Shortest Path. 153

SUs Secondary Users. 23

SV Sampled Values. 168

TM Traffic Matrix. 162

Abstract

Reinforcement learning is a branch of machine learning that enables machines to learn by trial and error. It is an experience-driven sequential learning process to achieve a particular goal. Recent advances in reinforcement learning have combined deep learning, which has led to the emergence of a new field called deep reinforcement learning (DRL). DRL algorithms have shown great success on various complex decision-making tasks that were earlier thought to be extremely difficult for a computer. Communication networks play a fundamental role in today's information age, where connectivity has become a basic commodity of life. They will play an even more critical role in the future, when *everything* from people, animals, wearable devices and cars to buildings and industries, will be connected. Providing connectivity on such a massive scale calls for an advanced set of solutions that can deal with complex, large-scale, and dynamic wireless and wired networks. DRL has the potential to meet these challenges due to its ability to learn from experience and adapt to the changing complex decision-making environment. Thus, we use DRL as a primary tool in this thesis and investigate one wireless and one wired technology.

Spectrum scarcity is one of the major issues for wireless communications, due to the limited availability of spectrum bands. A cognitive radio network (CRN) is one of the potential technologies that can overcome spectrum scarcity by efficiently utilising idle spectrum. In a CRN, cognitive users dynamically utilise idle spectrum with the requirement of no harmful interference to the licensed spectrum users. The first part of this thesis investigates CRN from the viewpoint of enhancing the process of dynamic spectrum access, and identifies three research objectives. First, a non-cooperative distributed CRN is considered where the goal is to achieve fairness of spectrum resources while avoiding collision. This is a highly challenging research problem due to the non-cooperating network users and non-stationary environment. A multi-agent DRL solution is proposed that significantly improves the network performance in terms of fairness of spectrum resources

and collision avoidance among all users of the network. Second, the task of local and cooperative spectrum sensing in a centralised CRN is investigated. For licensed users, time-inhomogeneous discrete and continuous activity models are considered. An offline DRL solution is proposed that significantly improves the detection of licensed spectrum users and idle spectrum utilisation. Third, prediction-based spectrum sensing in a centralised CRN is investigated. A k-step prediction framework with an offline DRL solution is proposed for both local and cooperative spectrum sensing. The proposed solution significantly improves the energy efficiency and transmission time of cognitive users.

For wired communications, a software-defined network (SDN) is a logically centralised network that provides various benefits over traditional distributed networks, such as on-demand resource allocation and easy reconfiguration. In an SDN, quality of service (QoS) routing is basic functionality that determines the path between a source and destination which fulfils the QoS requirements. It plays a vital role in various crucial network services, such as substation communication in a smart grid. The second part of this thesis investigates QoS routing in an SDN. Existing DRL solutions formulate routing as a continuous control problem using k-shortest paths, whereas we propose a discrete control QoS routing framework called deep Q-routing that does not use predefined k-shortest paths. Instead, the proposed framework learns the network topology and optimises QoS parameters while avoiding network loops and invalid actions. The effectiveness of the proposed framework is shown for both soft and hard QoS routing.