

An Exploration of Spiking Neural Networks and their use on Reinforcement Learning Tasks

by **Andrew William Rafe**

Thesis submitted in fulfilment of the requirements for the degree of

Master of Science (Research) in Computing Sciences

under the supervision of Dr. William Raffe and Dr. Jaime Garcia

University of Technology Sydney

Faculty of Engineering and Information Technology

School of Computer Science

May 2021

Certificate of Original Authorship

I, Andrew William Rafe declare that this thesis, is submitted in fulfilment of the requirements for the award of MSc (Res) in Computing Sciences, in the School of Computer Science at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Production Note:
Signature removed prior to publication.

11th May 2021

This thesis is dedicated to my mum who, in the opening weeks of this research, lost her long battle with breast cancer. You always taught me to pursue my interests in life and your immense commitment to your children has given me the opportunities to do just that.

Acknowledgements

I would like to take this opportunity to thank the people in my life for whom without them, the completion of this thesis would not have been possible.

Firstly, I would like to show my gratitude to my supervisors who have supported me throughout and have always been available for me to rack their brains, get their opinions on the direction of the research and provide me with every needed support to get this thesis completed. Dr. William Raffe, my primary supervisor, would always be happy to engage in the technical discussions I needed to be able to articulate my thought process, and to keep me on track. To Dr. Jaime Garcia, my co-supervisor, you would always be there to help me through my communication of the research and has therefore allowed me to excel at presenting my ideas to a wider skill-based audience. I would also like to thank you both for creating a wonderful environment in the Game Studio Research Lab and providing me with the opportunity to teach at the university during the conduct of this research, greatly widening the skill set that this degree has provided me.

Secondly, I would like to thank all the members, new and old, of the Game Studio Research Lab which I have had the benefit of watching grow during my time with them. They would always be available to provide support during stressful assessment periods and would drop everything to attend the various candidature assessment stages to provide their moral support. Although we were not able to be face to face for a large portion of my degree, the online and in person meetups that we had along the way were some of the highlights of my time during these years. So, thank you all and I look forward to my continued engagement with the Game Studio Research Lab in future.

Lastly, but certainly not least, I would like to thank my family. To my dad, Barry, my sister, Christine, and my brother, Michael, thank you for the continued support through these difficult years. Through our family gatherings, zoom catchups and constant family chats, you have all really pushed me to keep going and to produce the best possible work. You would always be interested in what I was currently working on and would happily endure my far too detailed recounts of what I was doing and what I had found.

For all of you, I would like to say thank you and, without you, I would not have been able to get through this period of my life and I will be forever grateful.

Publications

Rafe, A.W., Garcia, J.A. & Raffe, W.L., 2021, June. Exploration Of Encoding And Decoding Methods For Spiking Neural Networks On The Cart Pole And Lunar Lander Problems Using Evolutionary Training. *In 2021 IEEE Congress on Evolutionary Computation (CEC)*, pp. 498-505. IEEE. doi: 10.1109/CEC45853.2021.9504921.

Contents

1	Introduction	1
1.1	Motivation	2
1.2	Research Gaps and Aims	4
1.3	Research Questions	5
1.4	Objectives	6
1.5	Methodology Constraints	7
1.6	Out of Scope	8
1.7	Significance	8
1.8	Summary	10
2	Background	11
2.1	Reinforcement Learning (RL)	11
2.2	Artificial Neural Network (ANN)	13
2.3	Deep Q Learning	14
2.3.1	Back-Propagation	14
2.4	Spiking Neural Network (SNN)	15
2.4.1	Action Potential and Hodgkin-Huxley Model (H-H)	16
2.4.2	Izhikevich Model	18
2.5	Decoding Outputs into Action Selection	19
2.5.1	Rate Coding	20
2.5.2	Temporal Coding	21
2.6	Summary	22

3	Literature Review	23
3.1	Computational Efficiency of SNNs	23
3.2	Hand Crafted SNNs	24
3.3	Hebbian Based Learning	26
3.4	Back-propagation of Spiking Neural Networks	30
3.5	Evolutionary Algorithms applied to Spiking Neural Networks	31
3.6	Coding Strategies	33
3.7	Discussion and Open Questions	34
3.7.1	Encoding	34
3.7.2	Decoding	35
3.7.3	Noise	35
3.7.4	Stochastically Firing Neurons	35
3.7.5	Threshold Adaptation	36
3.7.6	Model Complexity	36
3.7.7	Domain Complexity	37
3.8	Summary	38
4	Methodology	39
4.1	Problem Descriptions	39
4.1.1	Cart Pole	39
4.1.2	Lunar Lander	41
4.2	Input Encoding	42
4.2.1	Binary Encoding	43
4.2.2	Double Encoding	44
4.3	State Exposure Period	46
4.4	Output Decoding	46
4.4.1	First To Fire (F2F)	47
4.4.2	Rate	48
4.4.3	F2F Reset and Rate Reset	48
4.5	Codebase	49

4.5.1	Open AI Gym	49
4.5.2	Spiking Neural Network	49
4.6	Summary	50
5	Evolutionary Experiments	51
5.1	Experiment 1 - Initial Experimentation	51
5.1.1	Description of Networks	52
5.1.2	Description of Genetic Algorithm	52
5.1.3	SNN Decoding Method	52
5.1.4	Results	53
5.2	Experiment 2 - Transformed Input Space	55
5.2.1	Input Transformation	55
5.2.2	Method	56
5.2.3	Results	57
5.3	Experiment 3 - Comparative Analysis	58
5.3.1	Input Encoder	59
5.3.2	Method	60
5.3.3	Results	62
5.3.4	Discussion	67
5.4	Experiment 4 - Encoding Method Comparisons	69
5.4.1	Other Methods of Encoding	69
5.4.2	Method	71
5.4.3	Results	71
5.4.4	Discussion	73
5.5	Summary	74
6	Hebbian Based Experiments	75
6.1	Experiment 5 - R-STDP	75
6.1.1	Method	76
6.1.2	Results	76
6.2	Experiment 6 - R-STDP Weight Constraining	77

6.2.1	Method	78
6.2.2	Results	79
6.3	Experiment 7 - EF-STDP	80
6.3.1	Method	80
6.3.2	Results	81
6.4	Experiment 8 - EF-STDP Search for Better Learning Rates	83
6.4.1	Method	84
6.4.2	Results	85
6.5	Summary	85
7	Conclusions and Future Work	87
7.1	Answers to Research Questions	90
7.1.1	RQ1: Are SNN structures suitable at solving RL problems?	90
7.1.2	RQ2: What network constraining and network coding meth- ods improve training in RL environments?	91
7.1.3	RQ3: Due to the inability to use gradient based algorithms for the training of SNNs, are non-evolutionary learning tech- niques effective at solving RL problems?	94
7.1.4	Research Questions Conclusions	95
7.2	Limitations	95
7.2.1	Technical Limitations	96
7.2.2	Algorithmic Limitations	96
7.3	Future Work	97

List of Figures

2.1	A visual description of a Markov Decision Process (Sutton and Barto 2018)	12
2.2	An example of a set of action potentials occurring according to the Hodgkin and Huxley membrane capacitance model.	17
2.3	An example of random input into an Izhikevich neuron model with the parameters from equation 2.3	19
4.1	A screen shot of the Cart Pole experiment.	41
4.2	A screen shot of the Lunar Lander environment.	42
5.1	The best fitness achieved each generation. The random mutations on the SNNs weights are identical to the mutations of the ANNs.	53
5.2	The best fitness achieved each generation. The random mutations on the SNNs weights were 10 times more than those for the ANN signified by the GW label. The exposure times of 5, 10, 20, 50 and 100 were tested with a first-to-fire (F2F) decoding method for the SNNs.	54
5.3	Average fitness over 10 separate executions of the algorithm for the best agent in each generation for each of the time exposures tested.	58
5.4	An example of the input encoders connection to the SNN. In this example, the SNN has no hidden layer as the Input Encoder layer are not made up of IZ neurons and simply act as a pre-processing step for state space information.	59

5.5	The generation to reach the goal for single layer networks across all exposure periods and decoding methods for the Cart Pole problem.	64
5.6	The generation to reach the goal formultilayer networks with 16 hidden neurons across all exposure periods and decoding methods for the Cart Pole problem.	65
5.7	The generation to reach the goal for multilayer networks with 32 hidden neurons across all exposure periods and decoding methods for the Cart Pole problem.	66
5.8	The number of random actions taken on average for each of the exposure periods tested and decoding methods used for single layer networks in the Cart Pole problem.	67
5.9	A comparison of direct encoding methods using a network with no hidden layers in the Cart Pole problem.	72
5.10	A comparison of probability encoding methods using a network with no hidden layers in the Cart Pole problem.	73
6.1	Average fitness per 100-episode blocks for networks in the Cart Pole problem using R-STDP for learning with exposure periods of 60.	77
6.2	Average fitness per 100-episode blocks for networks in the Cart Pole problem using R-STDP with Weight Constraining methods.	79
6.3	Average fitness of 5 separate trials over each learning cycle for each of the exposure periods tested.	81
6.4	The average final fitness for the exposure period and learning rate combinations after 100 learning cycles	84
6.5	The learning rate fitness peaks of the moving average.	85

List of Tables

4.1	The original state space that the environment produces	40
4.2	The transformed state space for Lunar Lander to remove negative inputs.	42
4.3	The transformed state space in order for the inputs to work with the SNNs	43
4.4	The transformed state space for Lunar Lander to remove negative inputs.	45
5.1	The Transformed State Space to Remove Negative Input Values. . .	56
5.2	Table of results for Cart Pole experiment showing the highest average fitness achieved and the generation it achieved the goal by each decoding method, exposure period and network structure with a goal fitness of greater than 195.0.	62
5.3	Table of results for Lunar Lander experiment showing the highest average fitness achieved and the generation it achieved the goal (if goal was achieved) by each decoding method, exposure period and network structure with a goal fitness of greater than 200.0.	63
5.4	Correlation Coefficients for Cart Pole problem between the generation that each trial reached completion and the exposure period used for that trial.	66

Abstract

Artificial neural networks have recently been the prominent architecture for reinforcement learning tasks. However, there is emerging evidence that spiking neural networks can perform just as well and can retain this performance across similar environments. Spiking neural networks are experiencing a surge in popularity due to their potential for large efficiency gains when compared to their traditional artificial neural network counterparts. Though, when attempting to replicate the successes of artificial neural networks, challenges are faced due to their vastly different architectures and therefore differing methods for training and optimisation. As spiking neural networks are considered more biologically plausible, methods of training inspired by natural learning have been proposed. These methods have been minimally applied to complex reinforcement learning domains, instead typically focusing on supervised learning problems. This thesis aims to explore the use of spiking neural networks in reinforcement learning domains. Methods of evolutionary and spike timing based training will be explored. Additionally, an in-depth analysis of different encoding and decoding methods is conducted. This research also addresses the trends in the effect of the time period that a state is exposed to a spiking neural network on the performance of the networks.