FUSION SPARSE AND SHAPING REWARD FUNCTION IN SOFT ACTOR-CRITIC DEEP REINFORCEMENT LEARNING FOR MOBILE ROBOT NAVIGATION

MOHAMAD HAFIZ BIN ABU BAKAR

A thesis submitted in
fulfillment of the requirement for the award of the
Degree of Master of Electrical Engineering

Faculty of Electrical and Electronic Engineering
Universiti Tun Hussein Onn Malaysia

JULY 2023

DEDICATION

I dedicated this work to Allah s.w.t for giving me good health during this study and thesis writing. This humble effort I dedicated to

My beloved mother and father,

Rashidah Binti Che Amat and Allahyarham Abu Bakar Bin Mustaffa,

For encouragement, support and pray for my success,

And this is for you, Mak and Abah.

My beloved family,

For giving support morale to completing this study.

My beloved supervisor,

Dr. Abu Ubaidah Bin Shamsudin,

For advice and support, always assist me in providing solutions during the study and difficult times.

My beloved friends and UTHM teammates

Always been together and appreciate all your advice.

Alhamdullilah. Thank you all.

ACKNOWLEDGEMENT

In the name of Allah, the Most Gracious and the Most Merciful

First and foremost, I would like to acknowledge the Almighty God for His benevolence and for granting me wisdom and perseverance not only during the time of research and writing of this thesis but also throughout my life.

I would like to extend my heartfelt appreciation and gratitude to my supervisor, Dr. Abu Ubaidah bin Shamsudin, for his sincere and invaluable intellectual guidance extended to me throughout the years of my postgraduate studies. My sincere appreciation goes to the Ministry of Education Malaysia and Universiti Tun Hussein Onn Malaysia for providing me with financial support. Additionally, I would like to express my gratitude to all of my friends for their unwavering support and for being there for me during many difficult moments.

Last but not least, special thanks to my beloved parents for their blessings and unflinching insistence, as they have always encouraged me to never stop achieving my goals in life.

ABSTRACT

Nowadays, the progress in autonomous robots is being driven by the advancements in new technologies, particularly Deep Reinforcement Learning (DRL). DRL facilitates the autonomous navigation of robots by enabling them to interact with their environment and navigate automatically. Achieving accurate navigation is crucial, and the utilization of Soft Actor-Critic Deep Reinforcement Learning (SAC DRL) offers the most effective solution based on the principles of Reinforcement Learning (RL). However, certain weaknesses in SAC DRL have been identified, particularly in the exploration process for accurate learning with faster maturity. To address this issue, this research has designed and developed a solution based on an appropriate reward function to guide the learning process. Several types of reward functions based on sparse and shaping rewards in the SAC method have been proposed in this research. These include the reward function with angle correction (RFAC), the reward function without angle correction (RFWAC), the reward function without sparse reward (RFWSR), and the reward function without sparse reward and angle correction (RFWSRAC). These reward functions aim to investigate the effectiveness of mobile robot navigation learning. Through a series of experiments, the results demonstrate that the fusion of sparse and shaping rewards in the SAC DRL facilitates successful navigation of the robot to the target position, while also enhancing accuracy and maturity. Specifically, the incorporation of sparse rewards in the reward function leads to a significant improvement. The system with the sparse reward achieves the lowest average error of 4.989%, outperforming the system without sparse rewards, which exhibits the highest average error of 99.252%.

ABSTRAK

Pada masa kini, perkembangan robot autonomus didorong oleh kemajuan dalam teknologi terbaru, terutama Deep Reinforcement Learning (DRL). DRL memfasilitasi navigasi autonomus robot dengan memungkinkannya berinteraksi dengan lingkungan dan navigasi secara otomatis. Mencapai navigasi yang akurat sangat penting, dan penggunaan Soft Actor-Critic Deep Reinforcement Learning (SAC DRL) menawarkan solusi paling efektif berdasarkan prinsip-prinsip Reinforcement Learning (RL). Namun, kelemahan tertentu dalam SAC DRL telah diidentifikasi, terutama dalam proses eksplorasi untuk pembelajaran yang akurat dengan kematangan yang lebih cepat. Untuk mengatasi isu ini, kajian ini telah merancang dan mengembangkan penyelesaian berdasarkan fungsi ganjaran yang sesuai untuk memandu proses pembelajaran. Beberapa jenis fungsi ganjaran berdasarkan ganjaran yang jarang dan penyesuaian dalam kaedah SAC telah dicadangkan dalam kajian ini. Ini termasuk fungsi ganjaran dengan pembetulan sudut (RFAC), fungsi ganjaran tanpa pembetulan sudut (RFWAC), fungsi ganjaran tanpa ganjaran yang jarang (RFWSR), dan fungsi ganjaran tanpa ganjaran yang jarang dan pembetulan sudut (RFWSRAC). Fungsi ganjaran ini bertujuan untuk mengkaji keberkesanan pembelajaran navigasi robot mudah alih. Melalui beberapa siri eksperimen, keputusan menunjukkan bahawa penggabungan ganjaran jarang dan pembentukan dalam SAC DRL memudahkan navigasi robot ke kedudukan sasaran dengan berjaya, sambil meningkatkan ketepatan dan kematangan. Secara khusus, penyerapan ganjaran jarang dalam fungsi ganjaran menghasilkan peningkatan yang signifikan. Sistem dengan ganjaran jarang mencapai kesalahan purata terendah sebanyak 4.989%, mengungguli sistem tanpa ganjaran jarang yang mempunyai kesalahan purata tertinggi sebanyak 99.252%.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

RL	–	Reinforcement Learning
ML	–	Machine Learning
AI	–	Artificial Intelligence
DRL	–	Deep Reinforcement Learning
SAC	–	Soft Actor Critic
DDPG	–	Deep Deterministic Policy Gradient
DQN	–	Deep Q-Network
SARSA	–	(State-Action-Reward-State-Action)
MDP	–	Markov Decision Process
RFAC	–	Reward Function with Angle Correction
RFWAC	–	Reward Function without Angle Correction
RFWSR	–	Reward Function without Sparse Reward
RFWSRAC	–	Reward Function without Sparse Reward and Angle Correction
UAV	–	Unmanned Aerial Vehicles
AUV	–	Autonomous Underwater Vehicle
ROS	–	Robotic Operating System
2D	–	Two-Dimension

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Background of the research

Nowadays, with the rapid evolution of the modern era, mobile robots have emerged as portable and robust technologies that are highly versatile and suitable for a wide range of situations. They excel in applications such as rescue and logistics operations, where they offer superior alternative solutions to tackle challenging conditions. As technology advances, continuous efforts are made to enhance mobile robots and provide greater convenience to humans. Artificial intelligence (AI) based technologies, especially those utilizing fuzzy concepts [1], [2], have emerged as a prominent trend in robot control. Many researchers have conducted numerous studies involving AI, specifically Reinforcement Learning (RL) [3], [4], and have introduced various technologies for the navigation process in mobile robot applications. These studies reflect the ongoing trend of implementing new technologies to enhance the capabilities of mobile robots using AI, particularly the use of Machine Learning (ML) [5]–[7].

ML is an effective branch of AI that is widely used to train robots without the need for constant human supervision [8]. RL enables robots to learn the next action based on their interactions with the environment, empowering them to operate and control autonomously. In recent times, various ML algorithms have emerged, including Q-Learning [9], [10], State-Action-Reward-State-Action (SARSA) [11], Deep Q-Network (DQN) [12], Deep Deterministic Policy Gradient (DDPG) [13], and Soft Actor-Critic (SAC) [14]. These algorithms are expected to continue making significant contributions in the future. The aim of this research is to implement a hybrid method called Deep Reinforcement Learning (DRL), which combines RL techniques

with neural networks. The application of this approach will enhance the intelligence and utility of mobile robots, including their potential for integration into autonomous vehicle systems in the future.

Completing seemingly simple tasks for a robot, like navigation and movement control, actually involves complex behaviors. To achieve this, a massive amount of data is needed, with high-dimensional state observations for each data point. The reason for collecting such a large amount of data is due to the use of On-Policy DRL. However, this approach is inefficient because it only utilizes each sample once for learning. To tackle the challenge of dealing with high-dimensional states, researchers have turned to Off-Policy DRL. This method involves reusing previously collected learning samples [15], which greatly reduces the reliance on millions of samples to master a seemingly simple task. However, Off-Policy DRL has its own issues, such as problems with non-linear neural network stability and convergence [16]. In this context, the robot's agent must learn in an unknown environment without a specific sample memory. It needs to maximize exploration during the learning process. To address these challenges related to exploration, the SAC algorithm is employed. The SAC algorithm aims to maximize entropy in the off-policy method, effectively addressing stability and convergence problems [14]. Furthermore, the SAC algorithm represents the state-of-the-art approach for continuous action, significantly improving the robot's movement performance, especially in terms of navigation accuracy.

In the field of RL research, the effectiveness and optimization of the system are crucial for ensuring its efficiency. Consequently, various factors, including the developed policy, can have an impact on SAC. Among these factors, the reward function plays a critical role in addressing complex task issues like navigation [17], [18]. The reward function is a vital component in determining the performance of the learning process based on the RL concept. By implementing an appropriate reward system, training performance can be enhanced, and the learning process can be expedited during environmental adaptation.

Among the various previous suggestions for improving systems that involve reward functions [17]–[19], the SAC algorithm stands out as a fundamental algorithm based on reward functions. In this research, the SAC algorithm will be utilized as the primary engine to control and determine the movement of robots, specifically focusing on mobile robots. The experiment will center around testing and enhancing performance using the reward function system. Simulation will be employed on a

platform to illustrate the hypothesis. Furthermore, this research aims to specifically address the highlighted issue by pursuing the following objectives:

1.Develop Sparse and Shaping reward functions in the SAC system for mobile robot navigation, serving as the main fundamental controller.

2. Evaluate the performance of the developed system through training and simulation processes.

The reward function has emerged as a vital element within the SAC system. Assessing the effectiveness of robots in process navigation poses a significant challenge, mainly due to the utilization of the reward function during robot training. The contribution of this research can help improve autonomous navigation (AV) technology based on AI, specifically by enhancing performance in terms of accuracy and increasing the maturity of short-term learning.

1.2 Problem statement

Autonomous mobile robots are widely used in various industries [20], [21]. A previous research [22] has found that one of the factors that affects the accuracy of autonomous robot movement is the complex representation of the observation space when processing environmental information as input data. Therefore, in this research, the focus is on improving navigation accuracy as a priority when developing methods to support the system. As a result, continuous action has been identified as the optimal solution to address the issue of movement accuracy. In recent years, a subfield called SAC has emerged within the latest DRL techniques. SAC combines RL and neural networks to facilitate the development of autonomous processes [14]. SAC is particularly known for its robustness and suitability for handling high-dimensional data observations. Additionally, it offers practical benefits as the system can utilize a replay buffer to enhance the learning process [14], [23].

Based on various previous studies [6] and [24], the SAC algorithm has emerged as the cutting-edge method for controlling autonomous robots, especially in tasks involving continuous action for navigation. However, the effectiveness of this system relies on the use of a suitable reward function that facilitates the learning process and motivates the robot to achieve its goals. To address this concern, this research aims to improve the learning performance of the SAC algorithm by enhancing the reward

function system, with a specific focus on improving navigation accuracy. On the other hand, previous research [17]–[19] has emphasized the pivotal role of the reward function in enhancing the overall system performance and advancing the learning capabilities of robots. Therefore, the research can enhance learning accuracy within a short period by employing an appropriate reward function system within the SAC algorithm.

1.3 Research aims

The aims of this research can be summarized as follows:

a) To develop a controller based on the SAC algorithm in order to operate autonomous mobile robot navigation using appropriate reward functions. This will be achieved by implementing a fusion of sparse and shaping reward functions.
b) To assist the learning efficiency and accuracy of the robot through system development.

1.4 Objectives

The objectives identified for implementation in this research are as follows:

a) To design and develop a Sparse and Shaping reward function in SAC DRL for mobile robot navigation.
b) To evaluate the performance of the developed method in terms of learning accuracy.

1.5 Scopes and limitations

In order to fulfil the stated objectives, the scope of this research will be divided into three stages:

i) Algorithm identification stage
   - Analysis of the Sparse and Shaping reward functions in SAC DRL for implementation in the control system.

ii) Training stage

- Train the agent using 1000 episodes for each reward function.
- Implement the SAC algorithm for a two-wheel differential drive in MATLAB and Simulink for 2D simulation.

iii) Simulation stage

- Evaluate the trained agent after the training process using MATLAB and Simulink.

To develop a fusion sparse and shaping reward function in SAC DRL for mobile robot navigation in this research, the following limitations have been identified:

a) Environmental factors, particularly weather conditions such as wind, rain, and unpredictable environments, are not considered since this research is based on a simulation platform.
b) In the development phase, the dynamic stability issues are not prioritized to ensure the stability of robot movement in achieving the objective.
c) The initial and target positions do not incorporate randomness because this research focuses on the development and performance of reward functions.
d) This research only utilizes the same map for the experiment since the main focus is on the reward function system.

This research includes a collaboration with The MathWorks Inc., focusing on the development of a SAC mobile robot navigation system using application simulation. MATLAB and Simulink are the primary software applications utilized for the simulation throughout the entire development process.

1.6 Outline of the thesis

This thesis proposes the utilization of fusion sparse and shaping reward functions in SAC DRL for mobile robot navigation. The primary objective of this research is to develop a simulation system that enables a novel approach to directly control mobile robot navigation without the need for supervision. This will be accomplished through the implementation of the SAC method with an appropriate reward function system.

Chapter 1 introduces the thesis and sets the context for the research. It provides an overview of the problem statement and the motivation behind the research. The chapter outlines the research objectives, highlighting the significance of developing a new approach for mobile robot navigation using fusion sparse and shaping reward functions in SAC DRL.

Chapter 2 undertakes an extensive review of the related works that have employed RL in the control of mobile robots. This comprehensive examination encompasses various topics such as AI, RL, DRL, SAC, mobile robots, and reward functions. By critically analyzing the existing literature, this chapter aims to provide a solid foundation and a comprehensive understanding of the research landscape.

Chapter 3 presents the methodology employed in this research, offering a detailed description of the procedural steps involved in the development of the advanced simulation system. The chapter elaborates on the design considerations, data collection methods, and the implementation process of the fusion sparse and shaping reward function system within the SAC DRL framework. It ensures a thorough understanding of the methodology followed, enabling reproducibility and credibility of the research.

Chapter 4 delves into the simulation results and conducts a comprehensive analysis of the research outcomes. The obtained results are meticulously evaluated, providing insights into the performance and effectiveness of the proposed approach for mobile robot navigation. The chapter presents statistical analyses, and visual representations to validate the system's performance and highlight its strengths and limitations.

Finally, Chapter 5 concludes by summarizing the achievements of the research and putting forth recommendations for future research endeavors.

CHAPTER 2

LITERATURE REVIEW

2.1 Theory

This research will utilize the relevant fundamentals that will be discussed in this topic. Controlling the robot to achieve the objectives through the implementation of SAC DRL is a subset of AI, serving as the primary control mechanism for this system. This approach aims to enhance the robot's intelligence and robustness in unpredictable environments. Table 2.1 illustrates the overall framework of DRL based on the underlying theories.

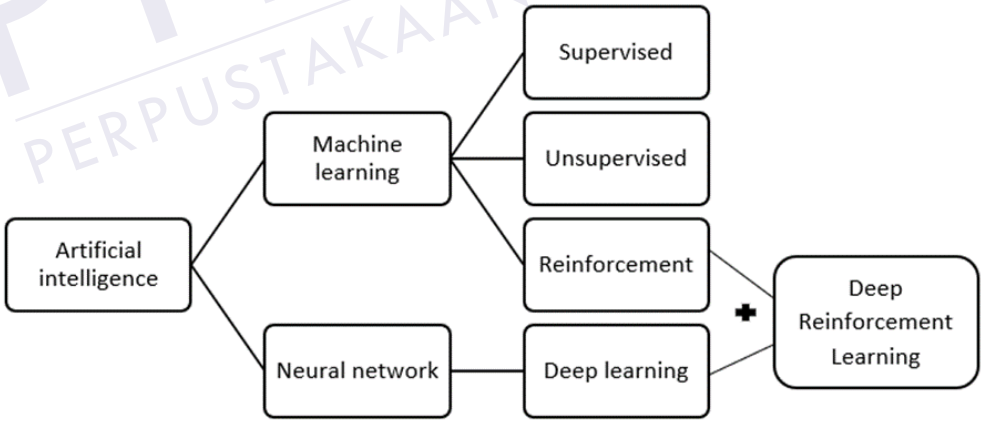

Figure 2.1: Block diagram for overall DRL.

2.1.1 Artificial Intelligence (AI)

AI is a field of science and technology that aims to replicate or simulate human intelligence in machines. Over the course of several decades, significant advancements have been made in AI, making it a key factor in measuring technological breakthroughs, especially in the robotics sector. In recent years, driven by the exponential growth of data and the availability of powerful computer hardware, AI has entered a new evolutionary stage with the development of related technologies, many of which are being implemented in robotics. One such technology is ML, which holds great potential for benefiting humanity. This research specifically focuses on the implementation of a hybrid algorithm that combines RL and deep learning (DL) to develop a system that enables autonomous control of robots without the need for manual intervention.

2.1.2 Machine Learning (ML)

ML is undoubtedly one of the most influential and powerful technologies in the modern era. This research begins with an introduction to the concepts of Machine Learning, which will be applied in mobile robots to enhance their intelligence and usefulness in the future. The objective is to cover fundamental ideas along with theoretical concepts, providing a comprehensive understanding.

ML can be categorized into three types: supervised, unsupervised, and RL [8], [25]. Each type of ML follows its own methodology and approach, but they all adhere to the same underlying principles and theories. This session will delve into the concept of ML, providing a thorough discussion and explanation. Table 2.1 summarizes the different types of ML.

Table 2.1: Summarise the different types of ML [8], [25], [26].

	Type of ML	Description	Algorithm of method	Application of algorithm
1	Supervised Learning	The method is applied to sample data. The target is to learn by the mapping concept (setting the rules) between a set of inputs and outputs.	Classification, Regression	image classification, object detection
2	Unsupervised Learning	In the training process for this method, only input data is provided without the guided sample. There is no labelled input to accomplish the goal. The model learns by observing and analysing. Furthermore, this method is more challenging to learn than Supervised Learning. The algorithm focuses less on the idea of determining the pattern and more on the unpredictable system.	clustering, association analysis	marketing automation
3	RL	The algorithm determines the target based on observations from interactions with the environment to take action that maximises reward or minimises risk. RL algorithms (as agents) constantly learn from their surroundings through error and update their experience until they reach their target.	Classification, Control	robotics, computer played games, autonomous car, data processing, aircraft control

2.1.3 Reinforcement Learning (RL)

RL is a popular ML technique in which an agent interacts with its environment through trial and error [26]. Essentially, RL can be compared to a baby learning to walk, improving through the learning curve and mistakes, thereby enhancing the learning process. Therefore, it is important to consider RL as a ML algorithm, alongside supervised and unsupervised learning, providing additional useful options depending on the system's development purpose.

In ML, RL is classified as a semi-supervised learning model. It enables an agent to act and interact with an environment to maximize total rewards. RL is typically modelled based on a Markov Decision Process (MDP) [26]. Figure 2.2

illustrates the basics of an RL system, where adjustment actions are made through the learning process to achieve the desired goal. Table 2.2 provides a description of the process and functions utilized in the RL system.

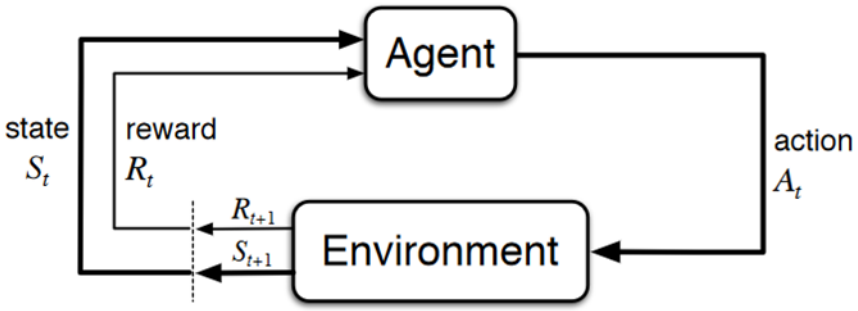

Figure 2.2: Based on modelled of RL in a Markov Decision Process [26].

Table 2.2: Describe of some important terms used in RL based on MDP.

	Process	Function
1	Agent	perform actions in an environment to gain reward
2	Environment	situation or scenario that an agent will be facing
3	Reward	an immediate return will give to an agent when performs for specific action or task
4	State	the current scenario or situation after returned response by the environment
5	Action	process or action will be taken based on environment and current stage

Most importantly, RL has several advantages: it can identify situations that require automated actions, discover optimal rewards, facilitate the agent's learning process to achieve those rewards, and enable the system to determine the most effective solutions or methods for obtaining rewards.

However, one of the disadvantages of RL is that it often requires a significant amount of time to solve complex tasks [27]. To address this issue, researchers have been increasingly exploring advanced technologies that combine two elements of AI, aiming to create more powerful and intelligent systems. This research focuses on the combination of Deep Neural Network (DNN) and RL, as it has shown promise in previous successful studies conducted by other researchers [6], [28], [29], [30], [31].

2.1.4 Deep Learning (DL)

DL is a subset of ML that aims to teach machines how to make decisions based on inputs and outputs, similar to the human brain. It utilizes neural networks with multiple layers of input to achieve the desired outcome. There are various types and structures of DL [32], including Boltzmann Machine, Deep Belief Network, Feedforward Deep Network, Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long-Short Term Memory (LSTM) network, and Generative Adversarial Network (GAN). The most popular structures used by researchers in DL are RNN and CNN. Figure 2.3 shows a block diagram of a neural network.

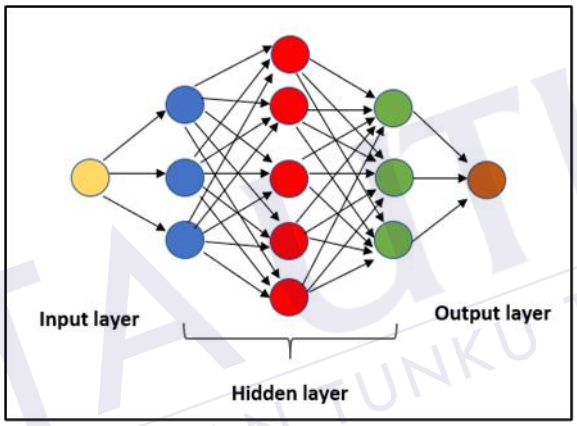

Figure 2.3: Block diagram of neural network.

DL is based on the concept of artificial neural networks, which employ algorithms to process large amounts of data and improve system efficiency. The advantage lies in the ability to effectively and efficiently process larger datasets. Figure 2.4 provides an illustration of the basic concept of DL.

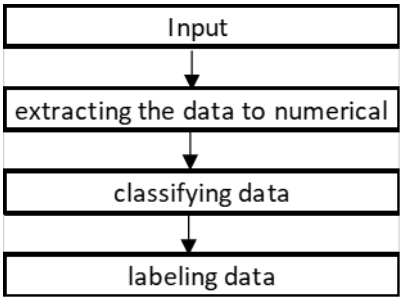

Figure 2.4: The basic concept in DL.

REFERENCES

[1] Xiaoyu Yang, M. Moallem and R. V. Patel, "A layered goal-oriented fuzzy motion planning strategy for mobile robot navigation," in IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 35, no. 6, pp. 1214-1224, Dec. 2005, doi: 10.1109/TSMCB.2005.850177.

[2] M. Faisal, M. Algabri, B. M. Abdelkader, H. Dhahri and M. M. Al Rahhal, "Human Expertise in Mobile Robot Navigation," in IEEE Access, vol. 6, pp. 1694-1705, 2018, doi: 10.1109/ACCESS.2017.2780082.

[3] Marc Peter Deisenroth, Gerhard Neumann and Jan Peters (2013), "A Survey on Policy Search for Robotics", Foundations and Trends® in Robotics: Vol. 2: No. 1–2, pp 1-142. http://dx.doi.org/10.1561/2300000021.

[4] A. S. Polydoros and L. Nalpantidis, "Survey of Model-Based Reinforcement Learning: Applications on Robotics," *J. Intell. Robot. Syst. Theory Appl.*, vol. 86, no. 2, pp. 153–173, 2017, doi: 10.1007/s10846-017-0468-y.

[5] J. Xiang, Q. Li, X. Dong and Z. Ren, "Continuous Control with Deep Reinforcement Learning for Mobile Robot Navigation," 2019 Chinese Automation Congress (CAC), Hangzhou, China, 2019, pp. 1501-1506, doi: 10.1109/CAC48633.2019.8996652.

[6] J. C. de Jesus, V. A. Kich, A. H. Kolling, R. B. Grando, M. A. de S. L. Cuadros, and D. F. T. Gamarra, "Soft Actor-Critic for Navigation of Mobile Robots," *J. Intell. Robot. Syst. Theory Appl.*, vol. 102, no. 2, pp. 1–11, Jun. 2021, doi: 10.1007/s10846-021-01367-5.

[7] G. Chen et al., "Robot Navigation with Map-Based Deep Reinforcement Learning," 2020 IEEE International Conference on Networking, Sensing and Control (ICNSC), Nanjing, China, 2020, pp. 1-6, doi: 10.1109/ICNSC48988.2020.9238090.

[8] Rabi Narayan Behera, "A Survey on Machine Learning: Concept, Algorithms and Applications," International Journal of Innovative Research in Computer

and Communication Engineering, vol. 5, no. 2, pp. 8198–8205, 2017, [Online]. Available: www.ijircce.com.

[9] S. I. A. Meerza, M. Islam and M. M. Uzzal, "Q-Learning Based Particle Swarm Optimization Algorithm for Optimal Path Planning of Swarm of Mobile Robots," 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), Dhaka, Bangladesh, 2019, pp. 1-5, doi: 10.1109/ICASERT.2019.8934450.

[10] X. Luo, Y. Gao, S. Huang, Y. Zhao and S. Zhang, "Modification of Q-learning to Adapt to the Randomness of Environment," 2019 International Conference on Control, Automation and Information Sciences (ICCAIS), Chengdu, China, 2019, pp. 1-4, doi: 10.1109/ICCAIS46528.2019.9074718.

[11] C. S. Arvind and J. Senthilnath, "Autonomous RL: Autonomous Vehicle Obstacle Avoidance in a Dynamic Environment using MLP-SARSA Reinforcement Learning," *2019 IEEE 5th Int. Conf. Mechatronics Syst. Robot. ICMSR 2019*, pp. 120–124, May 2019, doi: 10.1109/ICMSR.2019.8835462.

[12] T. S. Somasundaram, K. Panneerselvam, T. Bhuthapuri, H. Mahadevan and A. Jose, "Double Q–learning Agent for Othello Board Game," 2018 Tenth International Conference on Advanced Computing (ICoAC), Chennai, India, 2018, pp. 216-223, doi: 10.1109/ICoAC44903.2018.8939117.

[13] Y. Wang, J. Tong, T. -Y. Song and Z. -H. Wan, "Unmanned Surface Vehicle Course Tracking Control Based on Neural Network and Deep Deterministic Policy Gradient Algorithm," 2018 OCEANS - MTS/IEEE Kobe Techno-Oceans (OTO), Kobe, Japan, 2018, pp. 1-5, doi: 10.1109/OCEANSKOBE.2018.8559329.

[14] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor." PMLR, pp. 1861–1870, Jul. 03, 2018, Accessed: Feb. 14, 2023. [Online]. Available: https://proceedings.mlr.press/v80/haarnoja18b.html.

[15] H. Maei, C. Szepesvári, S. Bhatnagar, D. Precup, D. Silver, and R. S. Sutton, "Convergent Temporal-Difference Learning with Arbitrary Smooth Function Approximation," Advances in neural information processing systems, vol. 22, 2009.

[16] A. J. Snoswell, S. P. N. Singh, and N. Ye, "Revisiting Maximum Entropy Inverse Reinforcement Learning: New Perspectives and Algorithms," *2020*

IEEE Symp. Ser. Comput. Intell. SSCI 2020, pp. 241–249, Dec. 2020, doi: 10.1109/SSCI47803.2020.9308391.

[17] C. Wang, J. Wang, Y. Shen and X. Zhang, "Autonomous Navigation of UAVs in Large-Scale Complex Environments: A Deep Reinforcement Learning Approach," in IEEE Transactions on Vehicular Technology, vol. 68, no. 3, pp. 2124-2136, March 2019, doi: 10.1109/TVT.2018.2890773.

[18] W. Wang, Z. Wu, H. Luo, and B. Zhang, "Path Planning Method of Mobile Robot Using Improved Deep Reinforcement Learning," Journal of Electrical and Computer Engineering, vol. 2022, 1-7, 2022, doi: 10.1155/2022/5433988.

[19] J. Xie, Z. Shao, Y. Li, Y. Guan and J. Tan, "Deep Reinforcement Learning With Optimized Reward Functions for Robotic Trajectory Planning," in IEEE Access, vol. 7, pp. 105669-105679, 2019, doi: 10.1109/ACCESS.2019.2932257.

[20] M. Ben-Ari and F. Mondada, "Robots and Their Applications," Elements of Robotics, pp. 1–20, 2018, doi: 10.1007/978-3-319-62533-1_1.

[21] F. Rubio, F. Valero, and C. Llopis-Albert, "A review of mobile robots: Concepts, methods, theoretical framework, and applications," International Journal of Advanced Robotic Systems, vol. 16(2), 2019, doi: 10.1177/1729881419839596.

[22] J. Jhung, I. Bae, J. Moon, T. Kim, J. Kim and S. Kim, "End-to-End Steering Controller with CNN-based Closed-loop Feedback for Autonomous Vehicles," 2018 IEEE Intelligent Vehicles Symposium (IV), Changshu, China, 2018, pp. 617-622, doi: 10.1109/IVS.2018.8500440.

[23] C. Wang and K. Ross, "Boosting Soft Actor-Critic: Emphasizing Recent Experience without Forgetting the Past", 2019, doi: 10.48550/arxiv.1906.04009.

[24] M. H. Lee and J. Moon, "Deep Reinforcement Learning-based UAV Navigation and Control: A Soft Actor-Critic with Hindsight Experience Replay Approach," 2021, doi: 10.48550/arxiv.2106.01016.

[25] B. Mahesh, "Machine Learning Algorithms - A Review," International Journal of Science and Research, vol. 9, pp. 381–386, 2020, doi: 10.21275/ART20203995.

[26] R. S. Sutton and A. G. Barto, "*Reinforcement learning: An introduction*," MIT press, 2018.

[27] K. Arulkumaran, M. P. Deisenroth, M. Brundage and A. A. Bharath, "Deep Reinforcement Learning: A Brief Survey," in IEEE Signal Processing Magazine, vol. 34, no. 6, pp. 26-38, Nov. 2017, doi: 10.1109/MSP.2017.2743240.

[28] M. Wang, L. Wang and T. Yue, "An Application of Continuous Deep Reinforcement Learning Approach to Pursuit-Evasion Differential Game," 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chengdu, China, 2019, pp. 1150-1156, doi: 10.1109/ITNEC.2019.8729310.

[29] R. Takehara and T. Gonsalves, "Autonomous Car Parking System using Deep Reinforcement Learning," 2021 2nd International Conference on Innovative and Creative Information Technology (ICITech), Salatiga, Indonesia, 2021, pp. 85-89, doi: 10.1109/ICITech50181.2021.9590169.

[30] Q. Zhang, J. Lin, Q. Sha, B. He and G. Li, "Deep Interactive Reinforcement Learning for Path Following of Autonomous Underwater Vehicle," in IEEE Access, vol. 8, pp. 24258-24268, 2020, doi: 10.1109/ACCESS.2020.2970433.

[31] C. Wang, J. Wang, J. Wang and X. Zhang, "Deep-Reinforcement-Learning-Based Autonomous UAV Navigation With Sparse Rewards," in IEEE Internet of Things Journal, vol. 7, no. 7, pp. 6180-6190, July 2020, doi: 10.1109/JIOT.2020.2973193.

[32] A. Shrestha and A. Mahmood, "Review of Deep Learning Algorithms and Architectures," in IEEE Access, vol. 7, pp. 53040-53065, 2019, doi: 10.1109/ACCESS.2019.2912200.

[33] Sandeep Kumar Malu and Jharna Majumdar, “Kinematics, Localization and Control of Differential Drive Mobile Robot,” Global Journals of Research in Engineering, vol. 14, no. H1, pp. 1–7, Jan. 2014.

[34] T. Hellström, “Kinematics Equations for Differential Drive and Articulated Steering,” Department of Computing Science, Umeå University, 2011.

[35] Andrew. Y. Ng, “Shaping and policy search in reinforcement learning,” University of California, Berkeley, 2003.

[36] C. Stachniss, J. J. Leonard, and S. Thrun, “Simultaneous Localization and Mapping,” Springer Handbook of Robotics, pp. 1153–1176, 2016, doi: 10.1007/978-3-319-32552-1_46.

[37] V. Tompa, D. Hurgoiu, C. Neamtu and D. Popescu, "Remote control and

monitoring of an autonomous mobile robot," Proceedings of 2012 IEEE International Conference on Automation, Quality and Testing, Robotics, Cluj-Napoca, Romania, 2012, pp. 438-442, doi: 10.1109/AQTR.2012.6237750.

[38] T. Sai, D. Nakhaeinia, and B. Karasfi, "Application of Fuzzy Logic in Mobile Robot Navigation," Fuzzy Logic-Controls, Concepts, Theories and Applications, pp. 21-36, 2012.

[39] D. Babunski, J. Berisha, E. Zaev and X. Bajrami, "Application of Fuzzy Logic and PID Controller for Mobile Robot Navigation," 2020 9th Mediterranean Conference on Embedded Computing (MECO), Budva, Montenegro, 2020, pp. 1-4, doi: 10.1109/MECO49872.2020.9134317.

[40] N. Kumari, "International Journal of Computer Science and Mobile Computing Comparison of ANNs, Fuzzy Logic and Neuro-Fuzzy Integrated Approach for Diagnosis of Coronary Heart Disease: A Survey," *IJCSMC*, vol. 2, no. 6, pp. 216–224, 2013, Accessed: Feb. 14, 2023. [Online]. Available: www.ijcsmc.com.

[41] H. Batti, C. B. Jabeur and H. Seddik, "Fuzzy Logic Controller for Autonomous Mobile Robot Navigation," 2019 International Conference on Control, Automation and Diagnosis (ICCAD), Grenoble, France, 2019, pp. 1-6, doi: 10.1109/ICCAD46983.2019.9037922.

[42] A. Al-Mayyahi, W. Wang, and P. Birch, "Adaptive Neuro-Fuzzy Technique for Autonomous Ground Vehicle Navigation," Robotics, vol. 3, no. 4, pp. 349–370, Nov. 2014, doi: 10.3390/robotics3040349.

[43] J. K. Pothal and D. R. Parhi, "Navigation of multiple mobile robots in a highly clutter terrains using adaptive neuro-fuzzy inference system," Robotics and Autonomous Systems, vol. 72, pp. 48–58, 2015, doi: 10.1016/j.robot.2015.04.007.

[44] D. R. Parhi and P. K. Mohanty, "IWO-based adaptive neuro-fuzzy controller for mobile robot navigation in cluttered environments," The International Journal of Advanced Manufacturing Technology, vol. 83, no. 9–12, pp. 1607–1625, 2016, doi: 10.1007/s00170-015-7512-5.

[45] Z. Pezeshki and S. M. Mazinani, "Comparison of artificial neural networks, fuzzy logic and neuro fuzzy for predicting optimization of building thermal consumption: a survey," Artificial Intelligence Review, vol. 52, no. 1, pp. 495–525, 2019, doi: 10.1007/s10462-018-9630-6.

[46] M. K. Rath and B. B. V. L. Deepak, "PSO based system architecture for path planning of mobile robot in dynamic environment," 2015 Global Conference on Communication Technologies (GCCT), Thuckalay, India, 2015, pp. 797-801, doi: 10.1109/GCCT.2015.7342773.

[47] M. Li, W. Du, and F. Nian, “An adaptive particle swarm optimization algorithm based on directed weighted complex network,” Mathematical Problems in Engineering, vol. 2014, 2014, doi: 10.1155/2014/434972.

[48] M. J. Mohamed and M. K. Hamza, “Design PID Neural Network Controller for Trajectory Tracking of Differential Drive Mobile Robot Based on PSO,” Engineering and Technology Journal, vol. 37, no. 12A, pp. 574–583, 2019, doi: 10.30684/etj.37.12A.12.

[49] Aqeel-Ur-Rehman and C. Cai, "Autonomous Mobile Robot Obstacle Avoidance Using Fuzzy-PID Controller in Robot’s Varying Dynamics," 2020 39th Chinese Control Conference (CCC), Shenyang, China, 2020, pp. 2182-2186, doi: 10.23919/CCC50068.2020.9188467.

[50] G. Paragliola, A. Coronato, M. Naeem and G. De Pietro, "A Reinforcement Learning-Based Approach for the Risk Management of e-Health Environments: A Case Study," 2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Las Palmas de Gran Canaria, Spain, 2018, pp. 711-716, doi: 10.1109/SITIS.2018.00114.

[51] A. J. Almalki and P. Wocjan, "Exploration of Reinforcement Learning to Play Snake Game," 2019 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 2019, pp. 377-381, doi: 10.1109/CSCI49370.2019.00073.

[52] A. Jeerige, D. Bein and A. Verma, "Comparison of Deep Reinforcement Learning Approaches for Intelligent Game Playing," 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 2019, pp. 0366-0371, doi: 10.1109/CCWC.2019.8666545.

[53] V. Mnih *et al.*, “Human-level control through deep reinforcement learning,” *Nature*, vol. 518, no. 7540, pp. 529–533, 2015, doi: 10.1038/nature14236.

[54] Y. Duan, X. Chen, R. Houthooft, J. Schulman, and P. Abbeel, “Benchmarking deep reinforcement learning for continuous control,” 33rd International Conference on Machine Learning, pp. 1329-1338, 2016.

[55] B. R. Kiran et al., "Deep Reinforcement Learning for Autonomous Driving: A

Survey," in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 6, pp. 4909-4926, June 2022, doi: 10.1109/TITS.2021.3054625.

[56] T. T. Nguyen, N. D. Nguyen and S. Nahavandi, "Deep Reinforcement Learning for Multiagent Systems: A Review of Challenges, Solutions, and Applications," in IEEE Transactions on Cybernetics, vol. 50, no. 9, pp. 3826-3839, Sept. 2020, doi: 10.1109/TCYB.2020.2977374.

[57] C. Wu et al., "UAV Autonomous Target Search Based on Deep Reinforcement Learning in Complex Disaster Scene," in IEEE Access, vol. 7, pp. 117227-117245, 2019, doi: 10.1109/ACCESS.2019.2933002.

[58] K. Zhu and T. Zhang, "Deep reinforcement learning based mobile robot navigation: A review," in Tsinghua Science and Technology, vol. 26, no. 5, pp. 674-691, Oct. 2021, doi: 10.26599/TST.2021.9010012.

[59] M. Hillebrand, M. Lakhani, and R. Dumitrescu, “A design methodology for deep reinforcement learning in autonomous systems,” Procedia Manufacturing, vol. 52, pp. 266–271, 2020, doi: 10.1016/j.promfg.2020.11.044.

[60] Z. Yang, K. Merrick, L. Jin and H. A. Abbass, "Hierarchical Deep Reinforcement Learning for Continuous Action Control," in IEEE Transactions on Neural Networks and Learning Systems, vol. 29, no. 11, pp. 5174-5184, Nov. 2018, doi: 10.1109/TNNLS.2018.2805379.

[61] X. Yu, Y. Sun, X. Wang, and G. Zhang, “End-to-End AUV Motion Planning Method Based on Soft Actor-Critic,” Sensors 2021, vol. 21, no. 17, p. 5893, Sep. 2021, doi: 10.3390/s21175893.

[62] A. Hussein, E. Elyan, M. M. Gaber and C. Jayne, "Deep reward shaping from demonstrations," 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 2017, pp. 510-517, doi: 10.1109/IJCNN.2017.7965896.

[63] Y. Hu, Y. Hua, W. Liu and J. Zhu, "Reward Shaping Based Federated Reinforcement Learning," in IEEE Access, vol. 9, pp. 67259-67267, 2021, doi: 10.1109/ACCESS.2021.3074221.

[64] M. Riedmiller et al., “Learning by Playing Solving Sparse Reward Tasks from Scratch.” PMLR, pp. 4344–4353, Jul. 03, 2018, Accessed: Feb. 15, 2023. [Online]. Available: https://proceedings.mlr.press/v80/riedmiller18a.html.

[65] J. Hare and E. Vasilaki, “Dealing with Sparse Rewards in Reinforcement Learning,” Oct. 2019, doi: 10.48550/arxiv.1910.09281.

[66] A. Trott, S. Research, S. Zheng, C. Xiong, and R. Socher, "Keeping Your Distance: Solving Sparse Reward Tasks Using Self-Balancing Shaped Rewards," *Adv. Neural Inf. Process. Syst.*, vol. 32, 2019.

[67] A. Laud and G. DeJong, "The Influence of Reward on the Speed of Reinforcement Learning: An Analysis of Shaping," *Proceedings, Twent. Int. Conf. Mach. Learn.*, vol. 1, no. 1993, pp. 440–447, 2003.

[68] "Soft Actor-Critic Agents - MATLAB & Simulink." https://www.mathworks.com/help/reinforcement-learning/ug/sac-agents.html (accessed Feb. 15, 2023).

[69] Bharath Ramsundar and Reza Bosagh Zadeh "TensorFlow for Deep Learning," O'Reilly Media, Inc., 2018.

[70] A. D. Rasamoelina, F. Adjailia, and P. Sincak, "A Review of Activation Function for Artificial Neural Network," *SAMI 2020 - IEEE 18th World Symp. Appl. Mach. Intell. Informatics, Proc.*, pp. 281–286, Jan. 2020, doi: 10.1109/SAMI48414.2020.9108717.

[71] "Soft Actor-Critic Agents - MATLAB & Simulink." https://www.mathworks.com/help/reinforcement-learning/ug/sac-agents.html (accessed Feb. 14, 2023).

APPENDIX D

LIST OF PUBLICATIONS

1. M. H. Abu Bakar, A. U. Shamsudin and R. A. Rahim, "Simulation of Drone Controller using Reinforcement Learning AI with Hyperparameter Optimization," 2020 IEEE 10th International Conference on System Engineering and Technology (ICSET), Shah Alam, Malaysia, 2020, pp. 167-172, doi: 10.1109/ICSET51301.2020.9265381.
2. Mohamad Hafiz Abu Bakar, Abu Ubaidah bin Shamsudin, Ruzairi Abdul Rahim, Zubair Adil Soomro, & Andi Adrianshah. (2023). Comparison Method Q-Learning and SARSA for Simulation of Drone Controller using Reinforcement Learning. Journal of Advanced Research in Applied Sciences and Engineering Technology, 30(3), 69–78. https://doi.org/10.37934/araset.30.3.6978

APPENDIX E

VITA

The author was born on December 2, 1989, in Bukit Mertajam, Pulau Pinang, Malaysia. He attended SMK Munshi Sulaiman, a secondary school located in Batu Pahat, Johor, Malaysia. After completing his secondary education, he pursued a Diploma program in Electronic and Electrical Engineering at Politeknik Johor Bahru and graduated in 2009. Following his graduation, he gained valuable experience as a technician in the Department of TV Engineering at Sony EMCS in Bandar Baru Bangi, Selangor, Malaysia. In 2011, he enrolled at Universiti Tun Hussein Onn Malaysia in Parit Raja, Johor, Malaysia, where he successfully earned a Bachelor's degree in Electronic Engineering with a specialization in Mechatronics in 2014. Building upon his career achievements, he worked as a Test Engineer at Venture Technocom System Sdn Bhd in Johor Bahru, Johor, Malaysia, from 2016 to 2020. Driven by his passion for further knowledge and academic growth, he made the decision to join the Master's program in Electrical Engineering at Universiti Tun Hussein Onn Malaysia in December 2018.