DEEP REINFORCEMENT LEARNING FOR CONTROL

NURFARAH ANISAH BT MOHD YUSSOF

SCHOOL OF AEROSPACE ENGINEERING
UNIVERSITI SAINS MALAYSIA

2021

DEEP REINFORCEMENT LEARNING FOR CONTROL

by

NURFARAH ANISAH BT MOHD YUSSOF

Thesis submitted in fulfillment of the requirements for the Bachelor Degree of Engineering(Honours)(Aerospace Engineering)

June 2021

ENDORSEMENT

I, Nurfarah Anisah bt Mohd Yussof hereby declare that I have checked and revised the whole draft of dissertation as required by my supervisor.

(Signature of Student)

Date: 7 July 2021

Ye Zhou.

(Signature of Supervisor)

Name: Zhou Ye

Date : 7/7/2021

ENDORSEMENT

I, Nurfarah Anisah bt Mohd Yussof hereby declare that all corrections and comments made by the supervisor and examiner have been taken consideration and rectified accordingly.

(Signature of Student)

Date: 7 July 2021

Ye Zhou.

(Signature of Supervisor)

Name: Zhou Ye

Date : 7/7/2021

(Signature of Examiner)

Name: PM. Ir. Ts. Dr. Parvathy Rajendran

Date : 7 July 2021

DECLARATION

This thesis is the result of my own investigation except where otherwise stated and has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any other degree.

(Signature of Student)

Date: 7 July 2021

ACKNOWLEDGEMENTS

First and foremost, I would like to pay my special regards to my supervisor, Dr. Zhouye. She is an outstanding mentor who directed and motivated me on how to shape the task in the allotted period and do the right thing even when the situation becomes rough. This project's objectives would not have been fulfilled without her unwavering assistance. I wish to show my gratitude to Dr. Mohamad Aizat Abas for his amazing lecture in EMT 302 Mathematical Modelling in Engineering class, which gave an exposure to me the world of machine learning. I'd also wish to express my deepest gratitude to my friend, Siti Aisyah who co-supervised me and give me guidance and support to carry on this project. She is the one who always helps me with the Python Programming language as she is one of the experts in this field. Last but not least, I wish to thank my family and fellow friends whose assistance was a milestone in the completion of this project.

DEEP REINFORCEMENT LEARNING FOR CONTROL

ABSTRACT

Autonomous cars must be capable to operate in various conditions and learn from unforeseen scenarios. Driving a car with a human driver may be a challenging undertaking. As a result, autonomous driving seeks to reduce hazards in comparison to human drivers. Furthermore, autonomous driving is difficult in terms of the outcomes and safety judgments that must be taken. In this thesis work, a method using deep reinforcement learning to train a controller with proper driving behavior has been proposed. In essence, the method is to use a reward-based learning environment to watch how the agent makes decisions. Potential actions must be taken based on prior experiences using a trial and error process. However, determining the essential behavioral outputs for autonomous driving vehicle systems or selecting the optimal output features to learn from them is not easy. Deep Neural Networks were chosen as function estimators because of their capacity to solve the complexity of high-dimensional system issues. As a consequence, the agent is expected to have trained behaviors and navigation without crashing. The complete project is carried out in the CARLA simulator to determine how to operate in discrete action space using Deep Reinforcement Learning (DRL) algorithms. Gathering and evaluating a large amount of data is time and effort-intensive. Learning a model in a virtual environment might potentially fail to generalize to the actual world. As a result, the simulation environment makes it possible to collect massive training datasets. Improving learning driving policies can be adopted fast in the actual world. To generate the visual simulation in the simulator, the Python programming

language is employed. The improved algorithm will help encourage the real-world implementation of DRL in many autonomous driving applications.

PEMBELAJARAN PENEGUHAN MENDALAM UNTUK KAWALAN

ABSTRAK

Kereta autonomi mesti mampu beroperasi dalam pelbagai keadaan dan belajar dari senario yang tidak dijangka. Memandu kereta dengan pemandu manusia mungkin merupakan usaha yang mencabar. Hasilnya, pemanduan autonomi bertujuan untuk mengurangkan bahaya berbanding dengan pemandu manusia. Tambahan pula, pemanduan autonomi sukar dari segi hasil dan pertimbangan keselamatan yang mesti diambil. Dalam karya thesis ini, kaedah untuk menggunakan pembelajaran peneguhan mendalam untuk melatih pengawal dengan tingkah laku memandu yang betul telah diusulkan. Pada dasarnya, kaedahnya adalah dengan menggunakan persekitaran pembelajaran berasaskan ganjaran untuk melihat bagaimana ejen membuat keputusan. Tindakan berpotensi mesti diambil berdasarkan pengalaman sebelumnya menggunakan proses percubaan dan kesalahan. Namun, menentukan output tingkah laku yang penting untuk sistem kenderaan memandu yang autonomi atau memilih ciri output yang optimum untuk dipelajari. Rangkaian Saraf Dalam dipilih sebagai penganggar fungsi kerana kemampuan mereka untuk menyelesaikan kerumitan masalah sistem dimensi tinggi. Akibatnya, ejen tersebut diharapkan mempunyai tingkah laku dan navigasi terlatih tanpa terhempas. Projek lengkap dijalankan dalam simulator CARLA untuk menentukan bagaimana beroperasi di ruang tindakan diskrit menggunakan algoritma Pembelajaran Pengukuhan Dalam. Sebenarnya, mengumpulkan dan menilai sejumlah besar data memerlukan banyak masa dan usaha. Mempelajari model dalam persekitaran maya berpotensi gagal untuk menggeneralisasikan dunia sebenar. Hasilnya, persekitaran simulasi memungkinkan untuk mengumpulkan set data latihan secara besar-besaran.

Meningkatkan dasar memandu pembelajaran dapat diguna pakai dengan pantas di dunia sebenar. Untuk menghasilkan simulasi visual di simulator, bahasa pengaturcaraan Python digunakan. Algoritma yang ditingkatkan akan membantu mendorong pelaksanaan DRL di dunia nyata dalam banyak aplikasi pemanduan autonomi.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AI	: Artificial Intelligence
API	: Application Program Interface
CNN	: Convolutional Neural Network
CPU	: Central Processing Unit
DDAC	: Deep Deterministic Actor Critic
DNN	: Deep Neural Network
DQN	: Deep Q Network
DRL	: Deep Reinforcement Learning
FPS	: Frame Per Second
GPS	: Global Positioning System
GPU	: Graphics Processing Unit
LIDAR	: Light Detection and Ranging
LQR	: Linear Quadratic Regulator
MDP	: Markov Decision Process
ML	: Machine Learning
MSE	: Mean Square Error
RAM	: Random Access Memory
RL	: Reinforcement Learning
RLQR	: Robust Recursive Linear Quadratic Regulator
TD	: Temporal Difference

LIST OF SYMBOLS

s	: State
A_t	: Action at t
Q_t	: Quality at t
$Q_t(s', a')$	: Maximum expected cumulative reward for future pair (s', a')
$Q_t(s, a)$	: Maximum expected cumulative reward for considered pair (s, a)
R_t	: Reward at t, dependent, like S_t , on $A_{t-1}$and S_{t-1}
S_t	: State at t
R	: Set of possible rewards
S	: Set of all nonterminal states
$p(s', r\|s, a)$	: Probability of transitioning to state s' , with reward r, from s, a
r	: Environment
t	: Discrete-time step
v	: Velocity
α	: Action
γ	: Discount Rate Parameter

CHAPTER 1: INTRODUCTION

1.1 Overview

The area of Artificial Intelligence (AI) seeks to comprehend intelligent creatures (Drozdov, Kim and Lazebnik, 2011) where Machine learning (ML) is the process of creating a model based on sample data. Developed as a branch of Artificial Intelligence (AI) to replace the feed for manually constructing computer systems. Machine learning may be defined as the process of "automating automation". Instead of the supplied input and output, machine learning allows computers to design their programs. Machine learning, in other terms, is the process of translating data into programs. It is often referred to as training data for making predictions or choices.

Reinforcement learning is a type of machine learning that learns by experience. From its previous experiences, the computer will educate itself to focus on the prospective action that has to be performed. Agents may behave in the environment to achieve the goal by maximizing defined environmental rewards. We're aware of how our environment reacts to what we do while we're learning to drive a car or having a conversation, and we can affect what happens through our actions. Learning via contact is a fundamental notion that underpins almost all learning and knowledge theories.

Furthermore, reinforcement learning has emerged as one of the most active areas of research in computer learning, artificial intelligence, and neural network research.

Deep reinforcement learning algorithms are widely employed in a variety of applications, such as gaming consoles and simulation control agents. The trial and error approach is used to learn complicated topics by making mistakes and avoiding them in the future. This is a system in which an agent makes assumptions and acts in the environment in exchange for incentives.

However, this machine learning has limitations, such as the fact that it is not desirable to work on basic tasks and that too much reinforcement learning might lead to incorrect outcomes. A combination of reinforcement learning and deep learning is commonly employed in the solution of many reinforcement learning issues.

One application of deep reinforcement learning in continuous control problems is autonomous driving cars. The capacity to recognize its surroundings and drive safely without the assistance of a human driver applies to a self-driving automobile, also known as an autonomous vehicle(Rouse, no date). As a result, the focus of this work was on learning efficiency in continuous control tasks. DRL approaches will be employed in the simulated training of self-driving cars for this reason. Even with a human driver, driving a car may be considered a difficult activity; so, autonomous driving provides a challenge in terms of the performance and safety decisions that must be made.

1.2 Motivation and Problem Statement

Solving complex problems from unprocessed, high-dimensional, and sensory information has long been an issue in reinforcement learning. The problem is to guarantee that an autonomous automobile drives safely without colliding with other cars, people, or roadworks. A notion that demonstrates the agents' ability to generalize experience to new conditions has to be demonstrated.

To propose an environment-assisted autonomous driving system that can shorten the time spent learning after getting hit. The key challenge with this method is creating a model that will adhere to a particular time and restriction limitations. The management's objective is to increase learning performance by employing the DRL approach.

1.3 Objectives

The objectives of this study are:

1. To study the deep reinforcement learning framework in performing complex tasks.
2. To analyze and improve the algorithm in training data for autonomous car driving in multiple conditions.

1.4 Thesis Outline

This thesis consists of 5 chapters that the readers will get to know about the Deep Reinforcement Learning for an autonomous driving vehicle that has been done by me. First of all, the introduction section of the thesis is to provide the reader with an overview of the present situation and applications of machine learning in various disciplines. After reading the introduction, readers will have a general idea of what will be covered in the thesis's subsequent chapters.

Next, the background research and literature review will be introduced in Chapter 2. The various use of reinforcement learning will be introduced in this chapter. Machine Learning and Deep Learning, which are now the state of the art in practically every domain of autonomous agents, will be introduced. The purpose is to provide a high-level overview of the technical features of these algorithms using various diagrams and equations.

In addition, in Chapter 3, the basics of the Deep Reinforcement Learning (DRL) framework, were explained. Various approaches for training a DRL agent have been addressed in this area. In the instance of RL, the usage of policies in various contexts is discussed using various visuals and equations. The fourth chapter contains a variety of experiments and findings obtained using the Deep Reinforcement Learning method. It also displays graphs comparing various types of experimental outcomes. Chapter 5 finishes with a summary of the thesis, a discussion of the work's shortcomings, and a consideration of the future area of study.

CHAPTER 2: LITERATURE REVIEW

This chapter outlines the key studies that have been conducted that are essential to understand the type of simulations performed in the latter section of the thesis.

2.1 Machine Learning

Machine Learning has emerged as one of the most game-changing technical advances of the last decade as it is known as the greatest technique to automate the process of detecting patterns in data and then dealing with it to deliver particular tasks. ML is helping firms to accelerate digital transformation and usher in an era of automation. Machine learning algorithms, as previously stated, have the potential to improve themselves through training. Three main strategies are used to train ML algorithms which are supervised learning, unsupervised learning, and reinforcement learning will be introduced.

2.1.1 Supervised Learning

Supervised learning is one of the most fundamental forms of machine learning. The job is done by training the algorithm on labeled data to train a function that maps input data to output data based on an example input-output pair. When class identities are known, there are various strategies in supervised learning for dealing with unbalanced datasets, including re-sampling, under-sampling, and over-sampling techniques, which may be coupled with ensemble methods (Schaul *et al.*, 2016). The data must be appropriately labeled for this approach to operate and it is incredibly effective when being utilized in the correct conditions. In addition, supervised learning is also often known as a machine learning and artificial intelligence subcategory. The weights are adjusted until the model is well fitted as the input data is fed into the model. The cross-validation process is happening in this stage.

On the other hand, the training dataset is fairly close to the final dataset and it gives the algorithm with the labeled parameters necessary for the task in terms of features. The program then seeks correlations between the parameters provided, generating a cause and effect link between the variables in the dataset. After training, the algorithm understands how the data works and the relationship between the input and the output. This solution is then deployed for use with the final dataset, from which it learns in the same manner as it did with the training dataset. This means that supervised machine learning algorithms will continue to develop even after they have been implemented, uncovering new patterns and associations as they train themselves on new data.

2.1.2 Unsupervised Learning

Unsupervised machine learning may be the opposite of supervised learning as it has the benefit of working with unlabelled data. It allows the model to operate on its own to identify missing data patterns and information. No human input is necessary to make the dataset machine-readable that will allow the software to work on much bigger datasets. As unsupervised learning is lacking labels to deal with, it will lead to the formation of hidden structures. Anomaly detection, neural networks, and clustering are some of the examples that use this learning method. It will make it more flexible because of the formation of these hidden structures as the algorithms may adapt to the input by modifying hidden structures dynamically. Although unsupervised learning might be more unpredictable than other natural learning approaches, greater post-deployment development could be offered using this technique.

2.1.3 Reinforcement Learning

A branch of machine learning that studies how intelligent creatures should behave in a given environment to maximize the concept of cumulative reward is known as Reinforcement Learning (RL). It also one of three fundamental machine learning paradigms besides supervised and unsupervised learning. How humans learn from data in their daily lives is inspired directly through the reinforcement learning method. The use of a trial and error algorithm to improve itself and learn from scenarios could provide both favorable and unfavorable outcomes. The interpreter encourages the answer by rewarding the algorithm. The algorithm is compelled to repeat until a better result is found if unfavorable outcomes occur. Agents must generate efficient representations of the environment from high-dimensional sensory inputs and utilize them to generalize prior experience to new conditions to perform reinforcement learning successfully in scenarios approaching real-world complexity (Mnih *et al.*, 2015).

The Reinforcement Learning setting could be in two modes which are in online and offline modes that will be explained bellows :

- Online: In this mode, the learner receives constant feedback from the environment, resulting in more effective learning. The data becomes accessible in a sequential manner, and the agent's behavior is gradually updated based on the input.
- Offline: In this mode, the experience is gained before the learning. The agent learns how to act as a result. This is referred to as Batch Reinforcement Learning.

In addition, the general pros and cons of reinforcement learning are :

Advantages of using reinforcement learning :

- Help in solving very complex problems that cannot be solved by conventional techniques.
- Long-term results although it is very difficult to be achieved.
- The results will be closed to achieve perfection where the learning model would be very similar to learning human beings.
- The model can be corrected during the training process and the same error could be reduced.
- Useful when the only way to collect information about the environment to interact.

The disadvantages of Reinforcement Learning :

- Reinforcement Learning as a framework could be wrong in many different ways, but the quality that makes it useful.
- Too much Reinforcement Learning may lead to an overload of states which can diminish the results.
- Unsuitable for solving basic issues.

- Need lots of data for computation and it will work well in video games as it can be a replay and getting lots of data seem feasible.

2.2 Deep Reinforcement Learning

Deep Reinforcement Learning is a type of machine learning and artificial intelligence that allows machines to learn from their actions in the same manner that people do. The term "deep" may be defined as several layers of artificial neural networks that mimic the structure of human brains. It needs a big amount of training data as well as substantial computational resources. In terms of usefulness in real-world applications, it is still in its infancy.

The usage of deep reinforcement learning applications in machine learning has quickly grown. AWS DeepRacer is a well-known example of a DRL application for autonomous driving cars. AWS DeepRacer is an open learning system that allows users of all skill levels to study and experiment with reinforcement learning, as well as experiment with and develop autonomous driving applications. It includes a built-in camera as well as an onboard computer module. The compute module is inferred for it to proceed down the road. This AWS DeepRacer is made up of three parts: a console, a vehicle, and a league. The AWS DeepRacer interface provides developers with a service job that allows them to train and assess reinforcement learning models in a simulated autonomous driving environment. This advancement is excellent, particularly in terms of lowering the learning curve(Guide, no date). Although we know that reinforcement learning has tremendous potential for solving real-world issues, getting things done in the actual world may be costly and time-consuming.

On the other hand, the usage of DRL in-game consoles has been a success. AlphaGo is the first computer program to defeat a good human Go, player, as well as the first to defeat a Go World Champion, and is unquestionably the greatest Go player in history. Go is a Chinese board game that dates back over 3000 years. Because of its intricacy, winning this game takes a great deal of multi-tasking and strategic thinking. This computer software employs a search tree and a deep neural network method. The Go board is defined as the input and processing of a variety of network layers comprised of millions of neuron-like relationships. AlphaGo is made up of two parts: a tree search process and a convolutional network that controls the tree search procedure. Two policy networks and one value network are trained, for a total of three convolutional networks of two kinds. The AlphaGO method is presented in the diagram below.

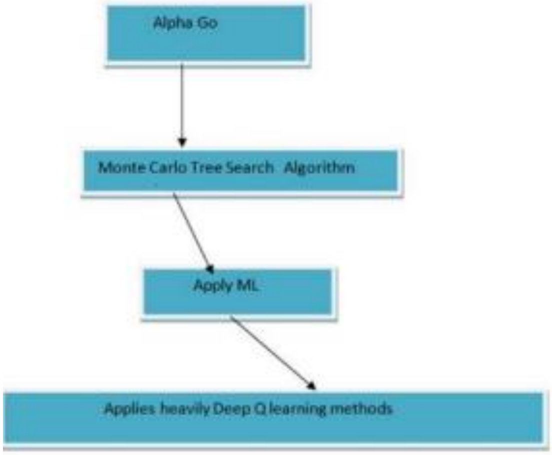
Alpha Go
Monte Carlo Tree Search Algorithm
Apply ML
Applies heavily Deep Q learning methods

Figure 2. 1: Alpha Go method (Deepmind, 2018)

Aside from that, a study is being performed to propose a hybrid vision-based route following controller for autonomous cars. The goal of this research is to create an autonomous car that will travel down the chosen path on a lane. This method employs a mix of DRL and the Robust Recursive Linear Quadratic Regulator (RLQR). CNN training using the reinforcement learning approach is capable of extracting the relevant