Vehicle Routing Problem Solving Using Reinforcement Learning

By

Shaneen Ara ,	ID:19201103138
Md.Mohiuddin Mostofa Kamal Akib ,	ID:17182103152
Esratjahan Sijana ,	ID:19201103099
MD.Tamim Hasan Opu ,	ID:18193103045
,	ID:

Submitted in partial fulfillment of the requirements of the degree of **Bachelor of Science** in

Computer Science and Engineering

Department of Computer Science and Engineering

Bangladesh University of Business and Technology

July 2023

Abstract

In this study, the Vehicle Routing Problem (VRP) is solved using reinforcement learning (RL) approaches. In order to service a group of consumers efficiently, the VRP involves identifying the shortest, most cost-effective, and fastest routes for a fleet of vehicles. With large-scale instances and dynamic situations, traditional methods for solving the VRP confront difficulties. A promising solution to this issue is provided by RL,a type of artificial intelligence. The suggested RL-based VRP solver learns to make wise routing decisions based on observable environmental information by utilising RL algorithms such as deep Q-networks (DQN) or proximal policy optimization (PPO). Experimental results show that the RL-based technique is effective, producing competitive or better solutions than conventional approaches, especially for complicated VRP situations.

Acknowledgement

We would like to express our heartfelt gratitude to the almighty Allah who offered our family and us kind care throughout this journey until the fulfillment of this research.Also, we express our sincere respect and gratitude to our course teacher **Khan Md. Hasib**, Assistant Professor, Department of Computer Science and Engineering, Bangladesh University of Business and Technology(BUBT). Without his guidance, this research work would not exist. We are grateful to him for his excellent guidence and for putting his utmost effort into developing this project. We owe him a lot for his assistance, encouragement, and guidance, which has shaped our mentality as a researcher.Finally, we are grateful to all our faculty members of the CSE department, BUBT, to make us compatible to complete this research work with the proper guidance and supports throughout the last four years.

Approval

This report **"Vehicle Routing Problem Solving Using Reinforcement Learning"** submitted by **Shaneen Ara-19201103138, Md. Mohiuddin Mostafa Kamal Akib-17182103152, Esratjahan Sijana-19201103099, MD.Tamim Hasan Opu-18193103045** Department of Computer Science and Engineering (CSE), Bangladesh University of Business and Technology (BUBT) under the supervision of **Khan Md. Hasib, Assistant Professor**, Department of Computer Science and Engineering (CSE) has been accepted as appeasement for the partial fruition of the requirement for the degree of Bachelor of Science (B.Sc.) in Computer Science and Engineering and endorsed as to its contents.

Supervisor:
MD.Shahiduzzaman
Assistant Professor
Department of Computer Science and Engineering (CSE)
Bangladesh University of Business and Technology (BUBT)
Mirpur-2, Dhaka-1216.

CourseTeacher:
Khan Md. Hasib
Assistant Professor
Department of Computer Science and Engineering (CSE)
Bangladesh University of Business and Technology (BUBT)
Mirpur-2, Dhaka-1216.

Contents

Introduction

1.1 Introduction

The Vehicle Routing Problem (VRP) is a well-known opti- misation problem with real-world applications in a variety of sectors, including distribution, logistics, and transportation. [1] It entails figuring out the best routes for a fleet of vehicles to take in order to serve a group of clients while minimising particular objectives, including the overall distance travelled, the cost paid, or the amount of time required. Improving op- erational effectiveness, cutting expenses, and raising customer happiness all depend on the VRP being solved effectively. [2] Heuristic algorithms or mathematical programming tech- niques are the foundation of conventional methods for solving the VRP. When dealing with complex problem situations or dynamic environments where consumer demands, traffic patterns, and delivery priorities might change quickly, these solutions frequently run into constraints. As a result, there is increasing interest in researching fresh ideas that can better manage the VRP's complexities and uncertainties.[3] Artificial intelligence's reinforcement learning (RL) sub- field has achieved outstanding results in resolving challenging decision-making issues across a wide range of applications. Based on feedback in the form of rewards or penalties, RL algorithms interact with the environment to develop the best rules. [4] Due to its ability to adjust and learn from the dynamic nature of the problem, this learning paradigm has the potential to address the issues raised by the VRP. [4] Investigating the use of RL approaches to resolve the VRP is the goal of this paper.

Our goal is to create an RL-based VRP solver that can learn to choose routes intelligently while taking into account variables like vehicle positions, client needs, road network data, and delivery limits. [5] We want to harness the potential of RL to find superior solutions to the VRP using state-of-the-art RL techniques, such as deep Q-networks (DQN) or proximal policy optimisation (PPO). [6] The ability of RL to manage dynamic settings and learn from experience is its key benefit for the VRP. Real-time routing decisions can be made by the RL agent while taking current information and alterations in the environment into account. [7]This versatility is especially useful in situations where new customer requirements come up, vehicles break down, or the delivery process is hampered by traffic.[8] In this paper, we'll outline the VRP's formulation as a reinforcement learning issue and go over the RL methods used to solve it. We will discuss the reward structure, the learning process, and the state and action representations. The performance and efficacy of the RL-based technique will be evaluated experimentally on benchmark VRP instances. The benefits and possibilities of RL in addressing the VRP will be illustrated through comparisons with conventional techniques. The topic of vehicle routing optimisation could be completely changed by the incorporation of reinforcement learning techniques into the VRP, opening the door to more effective and adaptive solutions. By utilising the power of RL, we can overcome the difficulties presented by the VRP and help numerous sectors optimise their logistics processes, cut costs, and improve customer service.[9]

1.2 Problem Statement

The problem is to solve the Vehicle Routing Problem (VRP) using Reinforcement Learning (RL) techniques. VRP involves optimizing the routing of a fleet of vehicles to serve a set of customers, aiming to minimize total travel distance, time, or other cost metrics, while satisfying various constraints such as vehicle capacity and customer time windows.

Traditional approaches to VRP often rely on heuristics, metaheuristics, or exact algorithms that require problem-specific knowledge or extensive computational resources. However, RL provides a promising alternative by learning optimal or near-optimal routing policies through interaction with the environment.

The problem is to design an RL-based framework that can effectively learn and improve routing policies for VRP. The framework should take into account the following challenges:

- **State Representation:** Designing a suitable state representation that captures the relevant information about the current state of the environment, including vehicle locations, customer demands, time windows, and other contextual factors.

- **Action Space:** Defining an appropriate action space that allows the model to choose the next action, such as selecting the next customer to visit or deciding to return to the depot.

- **Reward Design:** Formulating a reward function that provides informative feedback to the RL agent, incentivizing efficient and feasible routing decisions while considering multiple objectives like distance, time, and customer satisfaction.

- **Exploration-Exploitation Trade-off:** Balancing the exploration of new routing strategies with exploiting already learned knowledge to ensure the RL agent discovers optimal or near-optimal policies.

- **Scalability:** Developing an RL algorithm that can handle large-scale VRP instances with a significant number of vehicles, customers, and complex constraints, while maintaining computational efficiency.

- **Generalization:** Enabling the learned policies to generalize to unseen VRP instances or dynamically changing environments, allowing the model to adapt and make effective decisions in real-time scenarios.

The solution to this problem will contribute to developing intelligent routing algorithms for VRP, which can lead to more efficient resource allocation, reduced costs, and improved customer satisfaction in various industries such as logistics, transportation, and supply chain management.

1.3 Problem Background

The Vehicle Routing Problem (VRP) is a challenging optimization problem that involves determining optimal routes for a fleet of vehicles to serve a set of customers. VRP has practical applications in various industries, such as transportation, logistics, and delivery services. Traditional approaches to VRP often rely on heuristics or exact algorithms, which may struggle to find optimal or near-optimal solutions, especially for large-scale instances.

Reinforcement Learning (RL) is a branch of machine learning that enables agents to learn optimal decision-making policies through interaction with an environment. By applying RL techniques to VRP, we can train agents to make intelligent routing decisions without relying on explicit domain knowledge or pre-defined rules. RL agents learn from experience, receiving feedback in the form of rewards or penalties based on their actions.

Applying RL to VRP presents several challenges. Designing an appropriate state representation is crucial to capture the necessary information about the environment, including vehicle locations, customer demands, and time windows. The state representation should effectively encode the relevant information while maintaining computational efficiency.

Defining the action space is another challenge. The action space determines the choices available to the RL agent at each step, such as selecting the next customer to visit or deciding to return to the depot. The action space should be designed to allow for flexible and efficient routing decisions.

Reward design is a critical aspect of RL. Designing a suitable reward function that guides the agent towards optimal routing decisions while considering multiple objectives, such as minimizing travel distance and adhering to time constraints, requires careful consideration.

Balancing these objectives and providing informative feedback is essential for effective learning.

Exploration-exploitation trade-off is a challenge in RL-based VRP. Balancing exploration, which involves trying out new routing strategies, with exploitation, which involves leveraging already learned knowledge, is crucial for discovering optimal policies while making efficient decisions in real-time scenarios.

Scalability is another important consideration. VRP instances often involve a large number of vehicles, customers, and complex constraints. Developing scalable RL algorithms that can handle such instances efficiently is necessary for practical application.

Finally, generalization is important to ensure that the learned policies can adapt to unseen VRP instances or dynamic environments. The RL agent should be capable of making effective decisions and adapting to different routing scenarios, enhancing its robustness and real-world applicability.

Addressing these challenges in VRP using RL can lead to more efficient and effective routing strategies, reduced operational costs, improved resource utilization, and enhanced customer satisfaction in transportation and logistics industries.

1.4 Research Objectives

The research objectives for applying Reinforcement Learning (RL) techniques to solve the Vehicle Routing Problem (VRP) are as follows:

The objectives of our research work are as follows:

- Develop RL-based algorithms.
- Explore state representation techniques.
- Define action spaces.
- Formulate reward functions.

- Address exploration and exploitation.
- Handle large-scale instances.
- Evaluate performance and compare against benchmarks.

By achieving these research objectives, the aim is to advance the field of VRP by leveraging RL techniques to provide intelligent, adaptive, and efficient routing solutions. Such solutions have the potential to optimize resource allocation, reduce costs, improve customer service, and enhance operational efficiency in transportation and logistics industries.

1.5 Motivations

Adaptability: RL agents can learn and adapt their routing policies based on experience and observed consequences. They can dynamically adjust to changing conditions such as varying customer demands, traffic congestion, and road closures, ensuring optimal routing strategies in real-time.

Optimization: RL algorithms can efficiently explore the vast search space of possible vehicle routes, assignments, and scheduling decisions, aiming to find near-optimal solutions. By leveraging trial-and-error learning, RL agents can iteratively improve their decision-making and converge towards optimal routing policies.

Scalability: RL offers the potential for generalization and transfer learning. Once trained on a specific VRP instance, RL agents can generalize their learned policies to solve similar instances without requiring extensive retraining. This scalability feature reduces computational costs and accelerates the solution process for new VRP scenarios.

Dynamic Environments: RL is well-suited for handling the dynamic nature of the VRP. As customer demands and new orders arrive, RL agents can quickly adapt their routing strategies to efficiently allocate vehicles and satisfy time constraints, minimizing operational disruptions.

Sustainability: By optimizing vehicle routes and reducing unnecessary travel distances, RL-based routing strategies contribute to sustainability efforts. By minimizing fuel consumption and carbon emissions, RL can help businesses achieve more environmentally friendly and efficient transportation operations.

Learning from Experience: RL agents learn from trial and error, iteratively improving their routing policies over time. By exploring different actions and observing the resulting rewards, RL algorithms can discover efficient routing strategies, potentially outperforming traditional methods that rely on predefined rules or heuristics.

1.6 Flow of the Research

1.7 Significance of the Research

The research on applying Reinforcement Learning (RL) techniques to solve the Vehicle Routing Problem (VRP) holds immense significance. By leveraging RL, VRP solutions can achieve optimal or near-optimal routing strategies, leading to improved efficiency, reduced costs, and enhanced customer satisfaction. RL enables adaptive decision-making in real-time, allowing VRP solutions to dynamically respond to changing conditions and customer demands. Additionally, the research contributes to addressing scalability challenges, providing efficient solutions for large-scale VRP instances with complex constraints. The findings have practical

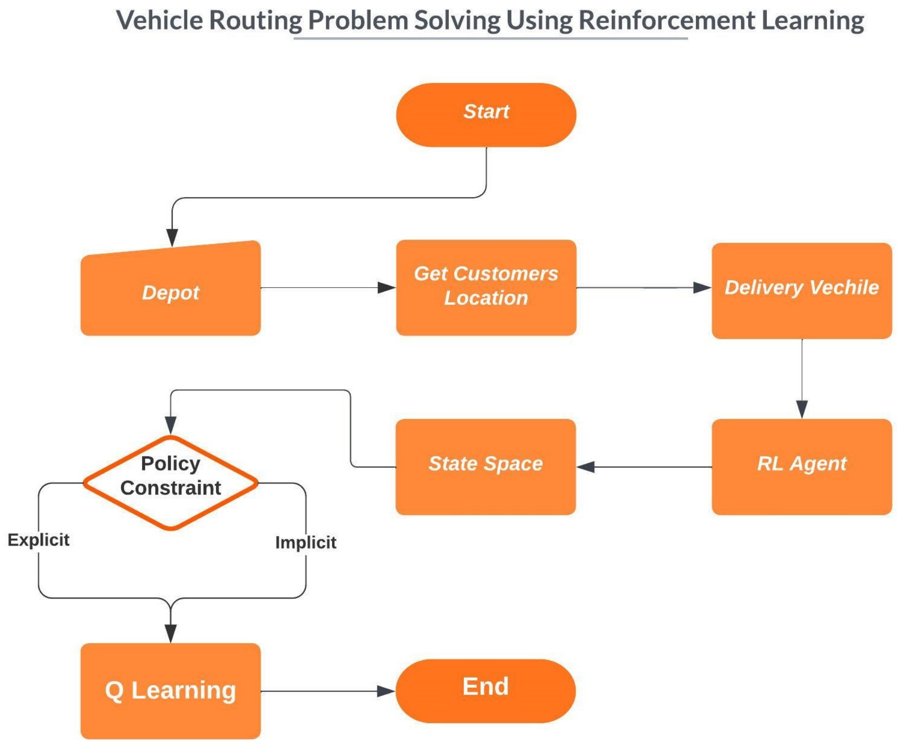


Figure 1.1. The figure illustrates the flow of the thesis work.

implications for industries such as logistics and transportation, enabling businesses to optimize their routing operations, improve resource allocation, and achieve competitive advantages. Overall, the research on VRP using RL has the potential to revolutionize the field of vehicle routing, making it more intelligent, adaptive, and efficient.

1.8 Research Contribution

The overall contribution of the research work are:

- **Optimal Routing Solutions:** The research contributes to the development of optimal or near-optimal routing solutions for VRP through RL. By training RL agents to learn efficient routing policies, the research enables businesses to minimize travel distance, time, or other cost metrics. This leads to improved resource utilization, reduced operational costs, and increased efficiency in vehicle routing.

- **Adaptive Decision-Making:** The research focuses on enabling adaptive decision-making in VRP through RL. RL agents learn to dynamically respond to changing conditions, such as varying customer demands, traffic congestion, or disruptions. This contribution ensures that VRP solutions can make real-time routing decisions, ensuring timely deliveries and effective resource allocation.

- **Scalability and Complexity Handling:** The research addresses the challenge of scalability and complex constraints in VRP using RL. By developing scalable RL algorithms, the research enables the efficient handling of large-scale VRP instances with a significant number of vehicles, customers, and intricate constraints. This contribution allows for practical implementation of RL-based VRP solutions in real-world scenarios.

- **Advancement in Routing Optimization:** The research contributes to the advancement of routing optimization techniques by leveraging RL in VRP. By exploring innovative state representations, action spaces, and reward functions, the research pushes the boundaries of traditional VRP algorithms. This contribution enhances the overall understanding and effectiveness of solving complex routing problems, benefiting industries such as logistics, transportation, and supply chain management.

1.9 Thesis Organization

This thesis is organized into several chapters to provide a comprehensive exploration of the research on solving the Vehicle Routing Problem (VRP) using Reinforcement Learning (RL) techniques.

Chapter 1 introduces the background and significance of the research, outlining the problem statement, objectives, and research contributions. It provides a brief overview of

VRP and RL, highlighting their relevance in the context of optimizing vehicle routing.

Chapter 2 presents a literature review, examining prior studies and approaches related to VRP and RL. This chapter synthesizes existing knowledge, identifies research gaps, and establishes the theoretical foundation for the proposed RL-based VRP solution.

Chapter 3 discusses the methodology employed in the research. It outlines the RL algorithms and techniques used, including state representation, action space design, reward formulation, and exploration-exploitation strategies. This chapter provides a detailed explanation of the experimental setup and data collection methods.

Chapter 4 presents the experimental results and analysis. It evaluates the performance of the RL-based VRP solution using various metrics, compares it with existing algorithms, and discusses the findings in relation to the research objectives. This chapter also explores the scalability and generalization capabilities of the proposed solution.

Chapter 5 discusses the implications and practical applications of the research findings. It highlights the potential impact of the RL-based VRP solution on industries such as logistics, transportation, and supply chain management. This chapter also addresses any limitations encountered during the research process and suggests avenues for future work.

Finally, Chapter 6 provides a concise summary of the research, reiterating the main contributions, and offers concluding remarks. It reflects on the research objectives and their fulfillment, and suggests possible directions for further exploration and development in the field of VRP using RL.

1.10 Summary

This chapter gives a thorough description of the issue that our study aims to address and explains how we were able to achieve our objectives. Additionally, it displays the broad procedures used to carry out our investigation.

Background

2.1 Introduction

In the domain of vehicle routing, traditional methods for solving the Vehicle Routing Problem (VRP) often involve manual planning and optimization, which can be time-consuming and prone to errors. The reliance on manual methods can lead to inefficiencies in route planning and increased operational costs. However, recent advancements in Reinforcement Learning (RL) techniques have shown promise in addressing these challenges.

Prior to this study, there has been limited research on the application of RL in VRP. The existing literature primarily focuses on heuristic or exact algorithms for solving VRP, leaving a gap in understanding the potential benefits of RL in this context. This research aims to bridge that gap by investigating the use of RL techniques to optimize vehicle routing, improve operational efficiency, and reduce costs.

2.2 Literature Review

In the suggested architecture, reinforcement learning is applied, and a single model is trained to identify nearly optimal responses to the VRP. The parameters of the model, which is a parameterized stochastic strategy, are optimized using the gradient policy approach. The trained model develops the solution in real-time by performing a series of subsequent actions while observing reward cues and adhering to feasibility constraints. The architecture is designed to address concerns with split delivery and capacitated VRP. [10]

Four-dimensional integer chromosomes are used to represent the VRPFIB issue, with each chromosome standing for a solution and including genes for vehicle, supplier, manufacturer, and customer IDs. To start the algorithmic search, a random population of chromosomes is created. The process of choosing parents uses a roulette wheel selection approach, with chromosomes with higher objective values having a larger chance of being picked. Order crossover is then used to apply crossover to the parent chromosomes, guaranteeing that no repeat alleles are served to any one client. Last but not least, the infant chromosomes go through swap mutation, which introduces minor genetic alterations by arbitrarily exchanging alleles between two genes. [11].

The General Variable Neighborhood Search (GVNS) metaheuristic, a local search strategy that investigates several neighborhoods to improve the result, is used in the suggested method in this study. The algorithm uses the Randomized Variable Neighborhood Descent (RVND) as the local search technique within the GVNS framework. The algorithm's main goal is to repeatedly improve the result by investigating various neighborhoods and making local adjustments. Regarding the parameter settings and the information utilized to test and characterize the method, the study report is deficient in specifics. Information on clients, depots, vehicle capabilities, and transportation charges are frequently included in test examples. [12]

The Firefly Algorithm (FA), which was first developed to solve continuous optimization issues, is the foundation of the suggested solution technique for the Electric-Parcel Consolidation Vehicle Routing Problem (E-PCVRP). The Firefly Algorithm based on Coordinates (FAC) is a modified version of the E-PCVRP since it is a discrete issue. The "Coordinates Related" (CR) encoding/decoding procedure is used by the FAC to resolve the difference. This method makes use of auxiliary vectors that hold the Cartesian coordinates of each node, allowing the FA's original movement equation to be applied directly. The FAC may more successfully explore the discrete search space of the E-PCVRP by incorporating the CR encoding/decoding process, making it a better optimization strategy for this particular issue. [13]

The objective of this study is to create heuristic methods to solve the Vehicle Routing Problem with Drones (VRPD). The main goal is to assess these strategies' performance using numerical tests on large-scale examples. The goal of the article is to show how integrating drones into last-mile logistics and delivery operations may have certain advantages. It may be assumed that the work concentrates on establishing solution methods based on heuristic approaches, which strive to produce near-optimal or high-quality answers utilizing efficient and effective procedures, even though the precise methodology and theory used to solve the VRPD are not described. [14]

In this research work, a novel parallelization technique for the genetic algorithm (GA)-based Traveling Salesman Problem (TSP) solution is proposed. The suggested approach is specially made to hasten the resolution of challenging vehicle routing problems (VRPs) in the context of cloud-based intelligent transportation systems. The solution efficiently offers routing data and related services for autonomous cars in vehicular clouds by parallelizing the GA. The approach uses three concurrent kernels, each performing GA-dependent operators, to address the time restrictions of intelligent transportation systems. Both multi-core and many-core processors may readily be used with this parallelization strategy, effectively leveraging their resources. The efficiency of the suggested strategy for parallelizing GAs on different processor architectures is demonstrated by experimental findings. [15]

This research report describes a similar study that uses a genetic algorithm to develop a waiting strategy for the vehicle routing issue with simultaneous pickup and delivery. In the context of online shopping and dynamic orders, the paper discusses the difficulties of real-world truck routing situations. The suggested waiting strategy takes into account current needs and seeks to streamline the routing procedure. The study proposes a rerouting indication for decision-making, even if the precise methods and theory are not given. The accuracy and performance of the evolutionary algorithm are evaluated by computational results when it comes to resolving difficult vehicle routing issues. [16]

The suggested approach uses an iterated local search algorithm that takes into account previous search data to direct the perturbation processes. The algorithm makes decisions on moves made during the local search and increases the likelihood of discovering better parts of the solution space by using the memory of earlier searches. Neighborhood restrictions, a straightforward heuristic for optimizing each route based on the Hamiltonian path's structure, and the use of elite solutions to focus the search on promising local optima are all included in the method. [17]

The suggested approach comprises solving the MDGVRP with competing objectives by using a multi-objective linear mathematical model. The traditional ACO algorithm is enhanced by the addition of a novel method for updating pheromone development. This novel method is applied by the IACO algorithm to update the pheromone and deliver improved results. Through the use of small- and large-scale networks that mimic real-world distribution issues, the model and method are proven. [9]

2.3 Problem Analysis

The manual method of data conversion and the lengthy processing time in the current system pose significant challenges. The reliance on manual data conversion leads to inefficiencies and potential errors in the process. Additionally, the time-consuming nature of the current system hinders timely completion of tasks. Furthermore, the presence of illegal registration methods in the land registration system in Bangladesh highlights the need for a more secure and reliable approach. The absence of prior work on the digitalization of land registration using the Hyperledger fabric framework further emphasizes the need for a comprehensive analysis of the problem. Therefore, this study aims to address these issues by proposing a digital version of the land registration system using the Hyperledger framework and blockchain technology, providing a secure and efficient solution to the existing problems in land registration processes.

2.4 Summary

The Vehicle Routing issue (VRP) is a difficult optimisation issue that entails choosing the best paths for a fleet of vehicles to take in order to serve a group of clients while taking into account a number of restrictions and minimising costs or distances. Heuristic methods or exact algorithms are frequently used in traditional approaches to VRP, which may have scaling issues or fail to discover the best answers. In recent years, there has been an increase in interest in using Reinforcement Learning (RL) approaches with VRP to learn effective routing strategies. Approaches based on RL have the capacity to manage complicated restrictions, adapt to changing contexts, and increase routing efficiency. This study intends to investigate the use of RL in VRP to produce optimal or nearly optimal solutions and improve operational performance in logistics and supply chain management.

Vehicle Routing Problem

3.1 Introduction

In this section, We describe all the topic of existing model of Vehicle Routing Problem. Here we show some figures and their problem statements.

3.2 Existing Methods for VRP

Reinforcement learning is used in the proposed framework, where a single model is trained to find close to ideal answers to the VRP. The model is a parameterized stochastic policy, and the parameters are optimized using a policy gradient technique.In real time, the trained model generates the solution as a series of subsequent actions while paying attention to reward cues and abiding by feasibility restrictions.The framework is built to deal with capacitated VRP issues and split delivery issues.

3.3 How to Experiment existing model?

To test the proposed framework, the researchers compare its performance with classical heuristics and existing software tools such as Google's OR-Tools.The researchers put the suggested framework to the test by evaluating how well it performs in comparison to conventional

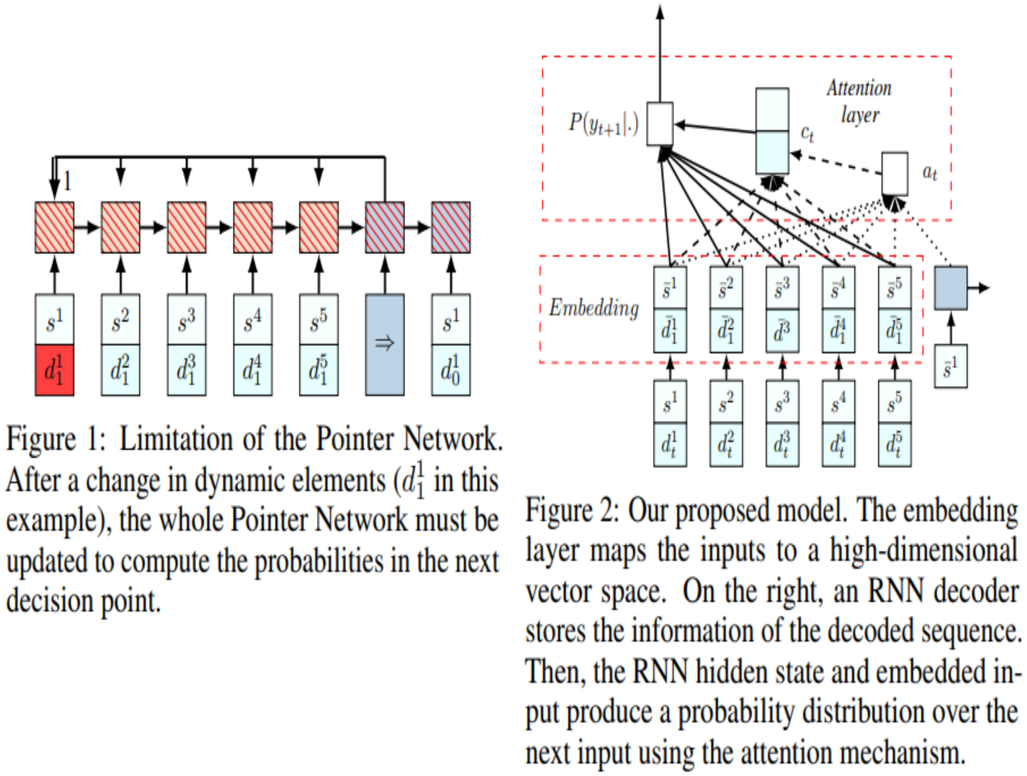


Figure 1: Limitation of the Pointer Network. After a change in dynamic elements (d_1^1 in this example), the whole Pointer Network must be updated to compute the probabilities in the next decision point.

Figure 2: Our proposed model. The embedding layer maps the inputs to a high-dimensional vector space. On the right, an RNN decoder stores the information of the decoded sequence. Then, the RNN hidden state and embedded input produce a probability distribution over the next input using the attention mechanism.

Figure 3.1. Existing Method

heuristics and current software programs like Google's OR-Tools.Performance is evaluated in terms of calculation time and solution quality. Additionally, the researchers investigate how split deliveries affect the caliber of solutions.The suggested framework is evaluated using medium- sized capacitated VRP instances.Using a cross validation methodology, the performance is assessed on examples that were not used for training.

3.4 Complications with the Present VRP models

There are several complications and challenges associated with the present Vehicle Routing Problem (VRP) models:

- **Scalability:** The scalability of VRP models is a significant challenge, especially for large-scale instances with a large number of vehicles, customers, and complex constraints. As the problem size increases, the computational complexity of finding optimal solutions grows exponentially, making it difficult to handle real-world VRP instances efficiently.

- **Dynamic Environments:** VRP models often assume a static environment, where customer demands, time windows, and road conditions remain constant. However, in real-world scenarios, these factors can change dynamically, requiring adaptive and real-time routing decisions. Existing models may struggle to address dynamic changes effectively, leading to suboptimal or infeasible solutions.

- **Complex Constraints:** VRP models need to consider various constraints, such as vehicle capacity, time windows, precedence constraints, and vehicle depots. Incorporating these constraints into the models while optimizing the routing decisions adds complexity and can make finding optimal solutions more challenging.

- **Solution Quality and Optimality:** Although existing VRP models provide solutions, ensuring their quality and optimality is a challenge. Heuristic and approximation algorithms are often used to find near-optimal solutions, but they may not guarantee the global optimum. Achieving high-quality solutions that are both efficient and optimal remains a complex task.

- **Integration with Real-time Data:** VRP models often rely on static or historical data, which may not capture real-time information such as traffic conditions or dynamic customer demands. Integrating real-time data into the models and adapting the routing decisions accordingly pose challenges in terms of data acquisition, processing, and decision-making.

3.5 Summary

In conclusion, research on applying Reinforcement Learning (RL) methods to solve the Vehicle Routing Problem (VRP) presents a viable path for enhancing routing tactics. Researchers want to train agents that can learn optimal or nearly optimal policies for vehicle routing, taking into account restrictions and minimising costs or distances, by utilising RL algorithms. Scalability, dynamic environments, complicated constraints, and solution optimality are issues that the application of RL in VRP addresses. In the logistics and transportation sectors, RL-based techniques have the potential to boost operational effectiveness, cut costs, and increase customer satisfaction. The scalability, adaptability, and real-time decision-making capabilities of RL in VRP need to be explored further, as well as a performance comparison with more conventional heuristic or precise algorithms.

Proposed Model, Testing, and Experimental Analysis

4.1 Introduction

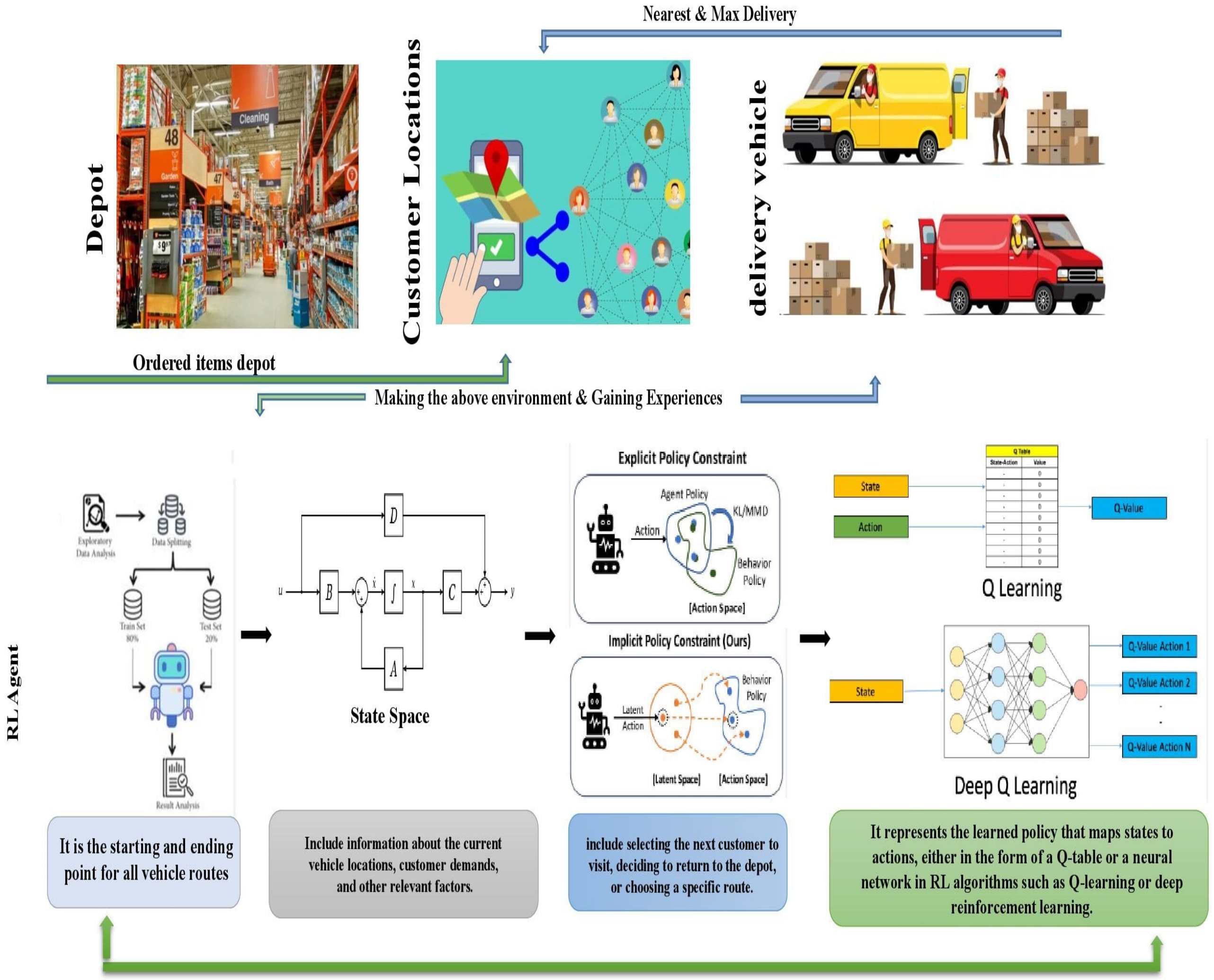
This section explains our proposed model workflow. Overall, this section describe how vehicle routing problem can be solve using reinforcement learning with Q learning.

4.2 Proposed Model Workflow

The proposed model for solving the Vehicle Routing Problem (VRP) using Reinforcement Learning (RL) with Q-Learning consists of three main components: the RL agent, the state representation, and the policy.

RL Agent: The RL agent is the core component of the model. It interacts with the environment, learns from experience, and makes routing decisions based on the learned policies. In the case of VRP, the RL agent represents the decision-maker responsible for selecting the next customer to visit, determining the sequence of visits, and deciding when to return to the depot. The RL agent uses Q-Learning, a popular RL algorithm, to update its Q-values based on the rewards received and optimize its routing policies over time.

State Representation: The state representation captures the relevant information about the current state of the VRP environment. It includes features such as vehicle locations, customer demands, time windows, and other contextual factors that influence routing decisions.



It is the starting and ending point for all vehicle routes

Include information about the current vehicle locations, customer demands, and other relevant factors.

include selecting the next customer to visit, deciding to return to the depot, or choosing a specific route.

It represents the learned policy that maps states to actions, either in the form of a Q-table or a neural network in RL algorithms such as Q-learning or deep reinforcement learning.

Figure 4.1. Proposed System of VRP using RL

The state representation should be designed to provide the necessary information for the RL agent to make informed routing choices.

Policy: The policy guides the RL agent's decision-making process by mapping states to actions. In the context of VRP, the policy determines which customer to visit next, the sequence of visits, and when to return to the depot. The RL agent's objective is to learn an optimal or near-optimal policy that maximizes the long-term cumulative reward, such as minimizing total distance traveled or meeting time constraints.

The proposed workflow for the VRP using RL with Q-Learning can be summarized as follows:

- Initialize the RL agent, including its Q-table or Q-network, with random values or pre-trained knowledge.
- Receive the current state of the environment, which includes information about vehicle locations, customer demands, and time windows.
- Initialize the RL agent, including its Q-table or Q-network, with random values or pre-trained knowledge.
- Receive the current state of the environment, which includes information about vehicle locations, customer demands, and time windows.
- Based on the current state, the RL agent selects an action (e.g., choosing the next customer to visit or deciding to return to the depot) using the policy derived from the Q-values.
- Execute the selected action in the environment and observe the resulting next state and the associated reward.
- Update the Q-values of the RL agent based on the observed reward and the transition from the current state to the next state, using the Q-Learning algorithm.
- Repeat steps 2-5 for multiple episodes or iterations, allowing the RL agent to learn and refine its policies.
- Evaluate the performance of the trained RL agent on new instances of the VRP, measuring metrics such as total distance traveled, adherence to time windows, and efficiency in resource utilization.
- Fine-tune the model parameters and iterate on the training process to further improve the RL agent's performance.

4.3 Working Strategy of Q-Learning in VRP using RL

The Q-Learning algorithm is a popular reinforcement learning technique used to solve complex sequential decision-making problems, including the Vehicle Routing Problem (VRP). Here is the working strategy of Q-Learning in VRP using RL:

4.3.1 Initialization:

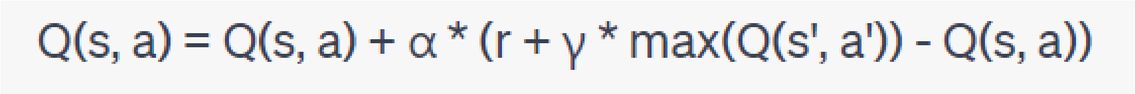
- Initialize a Q-table with dimensions representing states and actions. The states capture the VRP environment, such as vehicle locations, customer demands, and time windows. The actions correspond to selecting the next customer to visit, deciding the sequence of visits, and returning to the depot.
- Set initial Q-values for all state-action pairs in the Q-table

4.3.2 Exploration-Exploitation:

- Choose an exploration-exploitation strategy, such as epsilon-greedy, to balance between exploring new actions and exploiting the learned knowledge.
- During exploration, select a random action to explore the VRP environment and gather experience.
- During exploitation, select the action with the highest Q-value for the current state, based on the learned knowledge.

4.3.3 Update Q-Values:

- Execute the selected action in the environment and observe the next state and the associated reward.



Q(s, a) = Q(s, a) + α * (r + γ * max(Q(s', a')) - Q(s, a))

Figure 4.2. Updated Q value's rule

- Update the Q-value of the current state-action pair using the Q-Learning update rule:
- During exploitation, select the action with the highest Q-value for the current state, based on the learned knowledge. where:
  - Q(s, a) is the Q-value for state s and action a.
  - (alpha) is the learning rate, controlling the weight of the new information.
  - r is the reward received for taking action a in state s.
  - (gamma) is the discount factor, balancing the importance of immediate rewards and future rewards.
  - max(Q(s', a')) represents the maximum Q-value for the next state s' over all possible actions a'.

  Repeat Steps 3-4: Continue interacting with the environment, selecting actions, updating Q-values, and observing rewards until the termination condition is met (e.g., completing a fixed number of episodes or reaching a specific convergence criterion)

4.3.4 Training and Evaluation:

- Train the RL agent for multiple episodes or iterations, allowing it to learn and improve its policies.
- Evaluate the performance of the trained agent on new instances of the VRP, measuring metrics such as total distance traveled, adherence to time windows, and efficiency in resource utilization.

The Q-Learning strategy in VRP using RL aims to iteratively learn optimal or near-optimal routing policies by updating Q-values based on the rewards received during interactions with the VRP environment. Through exploration and exploitation, the RL agent learns to make efficient routing decisions, addressing the challenges of VRP and improving overall routing strategies.

4.4 Experimental Analysis

Choose a diverse set of VRP instances that represent real-world scenarios or synthetic problem instances. These instances should vary in terms of the number of vehicles, customers, customer demands, time windows, and other relevant factors. This ensures a comprehensive evaluation of the RL-based approach across different problem complexities.

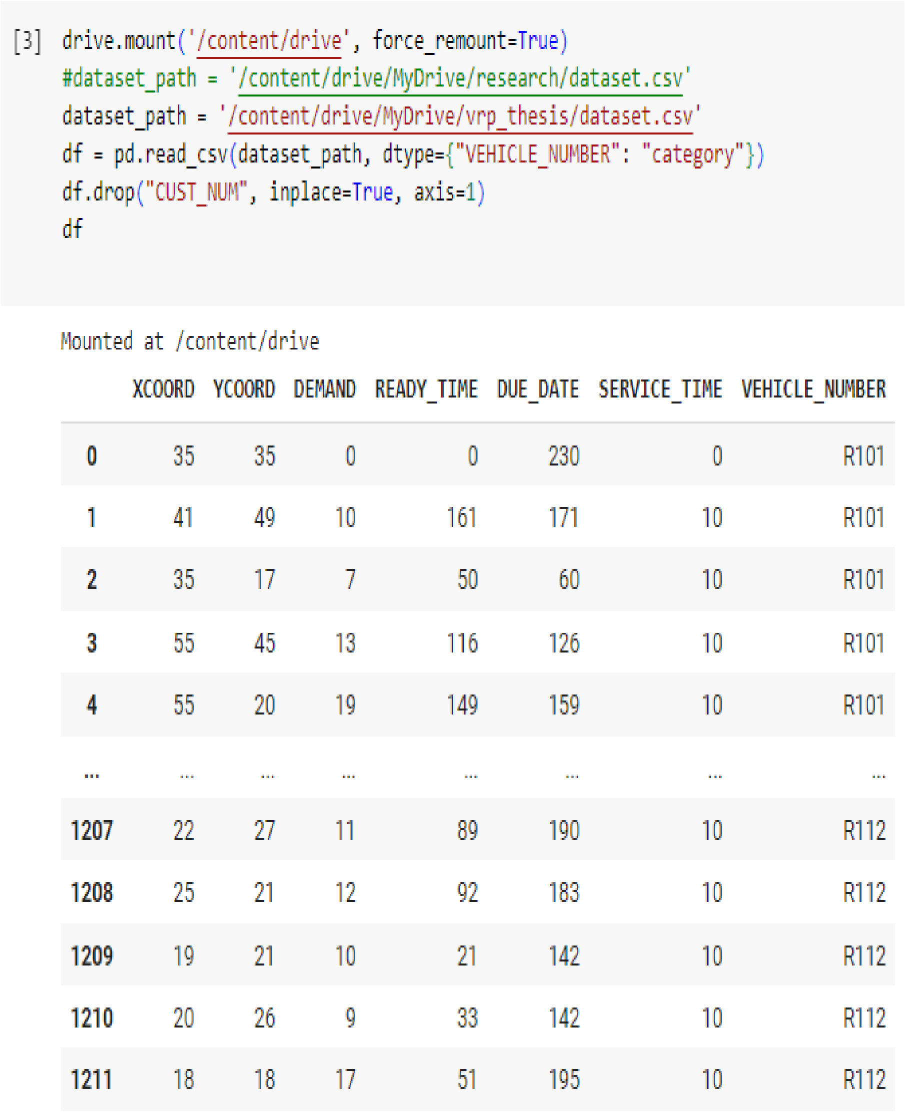
Implement the RL algorithms, such as Q-Learning, Deep Q-Networks (DQN), or Policy Gradient methods, to train the RL agent for solving VRP. Fine-tune the algorithm parameters, such as learning rate, exploration rate, and discount factor, to optimize the agent's performance.

Define appropriate performance metrics to evaluate the effectiveness of the RL-based approach. Common metrics for VRP include total distance traveled, percentage of on-time deliveries, total waiting time, and resource utilization. These metrics provide insights into the efficiency, effectiveness, and quality of the routing solutions obtained by the RL agent.

Conduct experiments by running the RL-based approach on the selected VRP instances and record the performance metrics. Repeat the experiments multiple times to account for randomness in the RL algorithms and obtain statistically significant results.

4.5 Summary

This section explains the architecture of the proposed model. The overall architecture uses the reinforcement algorithm and dataset trained using Q learning.



```
[3] drive.mount('/content/drive', force_remount=True)
    #dataset_path = '/content/drive/MyDrive/research/dataset.csv'
    dataset_path = '/content/drive/MyDrive/vrp_thesis/dataset.csv'
    df = pd.read_csv(dataset_path, dtype={"VEHICLE_NUMBER": "category"})
    df.drop("CUST_NUM", inplace=True, axis=1)
    df
```

```
Mounted at /content/drive
```

	XCOORD	YCOORD	DEMAND	READY_TIME	DUE_DATE	SERVICE_TIME	VEHICLE_NUMBER
0	35	35	0	0	230	0	R101
1	41	49	10	161	171	10	R101
2	35	17	7	50	60	10	R101
3	55	45	13	116	126	10	R101
4	55	20	19	149	159	10	R101
...	...	...	...	...	...	...	...
1207	22	27	11	89	190	10	R112
1208	25	21	12	92	183	10	R112
1209	19	21	10	21	142	10	R112
1210	20	26	9	33	142	10	R112
1211	18	18	17	51	195	10	R112

Figure 4.3. dataset

Standards, Constraints, Milestones

5.1 Standards

In the domain of Vehicle Routing Problem (VRP) using Reinforcement Learning (RL), the establishment of standards plays a crucial role in ensuring consistency, reliability, and interoperability. While specific standards for VRP using RL may still be emerging, it is important to adhere to existing standards and best practices in AI and operations research. This includes following algorithmic standards for transparent and reproducible RL model development, as well as adopting data standards for collecting, processing, and storing VRP data. Additionally, adherence to ethical standards, such as privacy protection, fairness, and transparency, is essential to promote responsible and ethical use of VRP solutions. By aligning with these standards, the VRP using RL community can promote robustness, reliability, and compatibility among different approaches and foster trust in the technology's applications.

5.2 Sustainability

Sustainability is a critical aspect to consider when applying Reinforcement Learning (RL) techniques to solve the Vehicle Routing Problem (VRP). VRP solutions using RL have the potential to contribute to sustainable practices in transportation and logistics. By optimizing vehicle routes, RL-based VRP solutions can minimize travel distances, reduce fuel consumption, and lower carbon emissions. This promotes efficient resource utilization and environmental conservation. Furthermore, RL algorithms can adapt to dynamic conditions, allowing for real-time adjustments to optimize routes based on traffic congestion, road conditions, or changing customer demands. Such adaptability leads to improved operational efficiency and reduced environmental impact. Incorporating sustainability considerations into the design and implementation of VRP using RL ensures a more responsible approach to transportation planning and contributes to a greener and more sustainable future.

5.3 Ethics

Ethics play a crucial role in the application of Vehicle Routing Problem (VRP) solutions. Ethical considerations include ensuring fairness, privacy, transparency, and accountability. It is essential to develop VRP systems that do not discriminate against specific individuals or groups, protect customer privacy and sensitive data, provide transparent explanations for routing decisions, and allow for human intervention and oversight. Adhering to ethical principles ensures the responsible and ethical use of VRP technologies, fosters trust among stakeholders, and promotes the well-being of individuals and society as a whole.

While applying Reinforcement Learning (RL) techniques to solve the Vehicle Routing Problem (VRP) offers promising opportunities, there are several challenges that researchers and practitioners need to address:

Exploration vs. Exploitation: Finding the right balance between exploration and exploitation is a key challenge in RL-based VRP. The RL agent needs to explore different routing options to discover optimal solutions, while also exploiting the learned knowledge to make efficient routing decisions. Striking the right balance can be challenging, especially in complex and dynamic VRP environments.

Scalability: Scaling RL-based VRP solutions to handle large-scale problem instances with a significant number of vehicles and customers remains a challenge. As the problem size increases, the computational complexity and memory requirements also grow. Developing scalable algorithms and approaches that can handle large and complex VRP instances is an ongoing challenge.

State Representation: Designing an effective state representation that captures the relevant information for VRP is crucial. The state representation should include factors such as vehicle locations, customer demands, time windows, and traffic conditions, among others. Determining the appropriate granularity and dimensionality of the state space is challenging and impacts the RL agent's ability to learn and generalize optimal routing policies.

Dynamic Environments: VRP often operates in dynamic environments where customer demands, road conditions, and other factors can change over time. Adapting RL-based VRP solutions to handle dynamic environments and incorporating real-time information poses a challenge. The RL agent needs to quickly adapt its routing decisions based on new information, requiring efficient algorithms and mechanisms for real-time decision-making.

Exploration of Policy Space: The policy space in VRP using RL is vast, and exploring it thoroughly to find optimal or near-optimal policies is a challenge. The RL agent needs to navigate a high-dimensional action space to discover efficient routing strategies. Efficient exploration methods, such as using function approximation techniques or combining RL with other optimization algorithms, are areas of active research to address this challenge.

Ethical Considerations: Ethical considerations, such as fairness, privacy, and transparency, need to be addressed in VRP using RL. Ensuring that routing decisions are fair and unbiased, protecting customer privacy, and providing transparent explanations for the decisions made by the RL agent are important challenges that need to be addressed to build trust and societal acceptance.

5.4 Summary

This chapter explains the standards, sustainability, impacts and ethics of the thesis work.

Conclusion

6.1 Introduction

In conclusion, applying Reinforcement Learning (RL) techniques to solve the Vehicle Routing Problem (VRP) holds significant promise for optimizing vehicle routing strategies. By leveraging RL algorithms, researchers aim to train RL agents that can learn optimal or near-optimal policies for efficient and cost-effective routing decisions. VRP using RL offers the potential to address challenges such as scalability, dynamic environments, complex constraints, and solution optimality. However, it also presents challenges in terms of exploration-exploitation trade-offs, scalability, state representation, adapting to dynamic environments, exploration of the policy space, and ethical considerations. Overcoming these challenges requires ongoing research, collaboration, and the development of scalable algorithms, efficient state representations, real-time data integration, and ethical frameworks. Despite the challenges, VRP using RL offers an innovative and promising approach to enhance operational performance, reduce costs, and improve resource utilization in transportation and logistics industries. With further advancements and interdisciplinary efforts, VRP using RL has the potential to revolutionize the field of vehicle routing and contribute to sustainable and efficient transportation systems.

6.2 Future Works and Limitations

Future Works: In the realm of Vehicle Routing Problem (VRP) using Reinforcement Learning (RL), future research can focus on advancing RL algorithms to improve their scalability, efficiency, and convergence properties. Exploring hybrid approaches that combine RL with other optimization techniques, addressing dynamic and stochastic VRP scenarios, and tackling multi-objective optimization are promising areas for further development. Additionally, real-world implementations and evaluations can provide valuable insights into the practicality and effectiveness of RL-based VRP solutions. **Limitations:**

It is important to acknowledge the limitations of VRP using RL approaches. Some limitations include:

- Computational Complexity: Solving large-scale VRP instances using RL can be computationally demanding and time-consuming. As the problem size increases, the training and inference times of RL models may become prohibitively high. Developing efficient algorithms to handle large and complex VRP instances is a challenge.
- Data Requirements: RL models require large amounts of training data to learn effective policies. Obtaining sufficient and diverse training data, especially for rare events or complex VRP scenarios, can be challenging. Data collection, cleaning, and labeling can also be time-consuming and resource-intensive.
- Generalization: RL models may struggle to generalize well to unseen or novel VRP instances. The learned policies may not perform optimally in situations that differ significantly from the training data. Enhancing the generalization capabilities of RL models in VRP is an ongoing research area.
- Ethical Considerations: RL-based VRP solutions must address ethical considerations, such as fairness, privacy, and transparency. Ensuring that the decisions made by RL agents align with ethical standards and comply with legal regulations is essential.

Developing robust ethical frameworks and guidelines for RL-based VRP is a critical aspect.

Addressing these limitations and exploring future research directions will pave the way for more effective and practical VRP solutions using RL, making significant contributions to the field of transportation and logistics optimization.

References

[1] Shahed Ahamed, Moontaha Siddika, Saiful Islam, SadiaSaima Anika, Anika Anjum, and Milon Biswas. Bps: Blockchain based decentralized secure and versatile light payment system. *Asian Journal of Research in Computer Science*, pages 12–20, 2021.

[2] Rami Ammourah and Alireza Talebpour. Deep reinforcement learning approach for automated vehicle mandatory lane changing. *Transportation research record*, 2677(2):712–724, 2023.

[3] Lu Duan, Yang Zhan, Haoyuan Hu, Yu Gong, Jiangwen Wei, Xiaodong Zhang, and Yinghui Xu. Efficiently solving the practical vehicle routing problem: A novel joint learning approach. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 3054–3063, 2020.

[4] Md. Jobaer Hossain, Md. Anwar Hussen Wadud, Anichur Rahman, Jannatul Ferdous, Md Shahin Alam, T. M. Amir Ul Haque Bhuiyan, and M. Firoz Mridha. A secured patient's online data monitoring through blockchain: An intelligent way to store lifetime medical records. In *2021 International Conference on Science Contemporary Technologies (ICSCT)*, pages 1–6, 2021.

[5] Suresh Nanda Kumar and Ramasamy Panneerselvam. A survey on the vehicle routing problem and its variants. 2012.

[6] Sm Al-Amin, Shipra Rani Sharkar, M Shamim Kaiser, and Milon Biswas. Towards a blockchain-based supply chain management for e-agro business system. In *Proceedings of*

International Conference on Trends in Computational and Cognitive Engineering, pages 329–339. Springer, 2021.

[7] Md. Jobaer Hossain, Md. Anwar Hussen Wadud, and Md. Alamin. Hdm-chain: A secure blockchain-based healthcare data management framework to ensure privacy and security in the health unit. In *2021 5th International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)*, 2021.

[8] Chaug-Ing Hsu, Sheng-Feng Hung, and Hui-Chieh Li. Vehicle routing problem with time-windows for perishable food delivery. *Journal of food engineering*, 80(2):465–475, 2007.

[9] Yongbo Li, Hamed Soleimani, and Mostafa Zohal. An improved ant colony optimization algorithm for the multi-depot green vehicle routing problem with multiple objectives. *Journal of cleaner production*, 227:1161–1172, 2019.

[10] Mohammadreza Nazari, Afshin Oroojlooy, Lawrence Snyder, and Martin Takác. Reinforcement learning for solving the vehicle routing problem. *Advances in neural information processing systems*, 31, 2018.

[11] Junayed Pasha, Maxim A Dulebenets, Masoud Kavoosi, Olumide F Abioye, Hui Wang, and Weihong Guo. An optimization model and solution algorithms for the vehicle routing problem with a "factory-in-a-box". *Ieee Access*, 8:134743–134763, 2020.

[12] SN Bezerra, SR de Souza, and MJF Souza. A gvns algorithm for solving the multi-depot vehicle routing problem. electron. notes discrete math. 66, 167–174 (2018). In *5th International Conference on Variable Neighborhood Search.*

[13] Dimitra Trachanatzi, Manousos Rigakis, Magdalene Marinaki, and Yannis Marinakis. A firefly algorithm for the environmental prize-collecting vehicle routing problem. *Swarm and Evolutionary Computation*, 57:100712, 2020.

[14] Daniel Schermer, Mahdi Moeini, and Oliver Wendt. Algorithms for solving the vehicle routing problem with drones. In *Intelligent Information and Database Systems: 10th Asian Conference, ACIIDS 2018, Dong Hoi City, Vietnam, March 19-21, 2018, Proceedings, Part I 10*, pages 352–361. Springer, 2018.

[15] Mahdi Abbasi, Milad Rafiee, Mohammad R Khosravi, Alireza Jolfaei, Varun G Menon, and Javad Mokhtari Koushyar. An efficient parallel genetic algorithm solution for vehicle routing problem in cloud implementation of the intelligent transportation systems. *Journal of cloud Computing*, 9:1–14, 2020.

[16] Hyungbin Park, Dongmin Son, Bonwoo Koo, and Bongju Jeong. Waiting strategy for the vehicle routing problem with simultaneous pickup and delivery using genetic algorithm. *Expert Systems with Applications*, 165:113959, 2021.

[17] José Brandão. A memory-based iterated local search algorithm for the multi-depot open vehicle routing problem. *European Journal of Operational Research*, 284(2):559–571, 2020.