RLC: A REINFORCEMENT LEARNING BASED CHARGING SCHEME FOR BATTERY SWAP STATIONS

by

Yutao Xu

Submitted in partial fulfillment of the requirements
for the degree of Master of Computer Science

at

Dalhousie University
Halifax, Nova Scotia
June 2023

To my parents, for the infinite love and support you have always provided.

Table of Contents

List of Tables

List of Figures

Abstract

Battery Swapping Station (BSS) is emerging as a promising solution to the prevalent issue of range anxiety among Electric Vehicle (EV) users. Typically, BSS replaces the drained battery of an incoming EV with a fully charged one. In this thesis, we propose a cutting-edge battery charging and swapping approach for BSS, termed Reinforcement Learning-based Charging (RLC). This innovative strategy enables the provision of partially charged batteries to EVs with lower energy requirements while simultaneously minimizing the overall energy expenditure of BSS. Technically, RLC employs an ensemble learning-based forecasting module to predict the electricity demand pertaining to EV battery swapping. Furthermore, it utilizes Deep Deterministic Policy Gradient (DDPG) to strategize the battery charging process within BSS. Specifically, the predicted electricity demand is fed into the DDPG agent, enabling it to adapt to the changing patterns of EV arrivals. Our experimental results indicate that RLC outperforms the baseline charging schemes in terms of overall electricity cost, average SOC discrepancy rate, and battery service rate. Our future work will focus on incorporating more real-life elements, such as dynamic electricity price and battery degradation, to further refine the proposed learning-based charging scheme.

List of Abbreviations

EV Electric Vehicle

BSS Battery Swap Station

MDP Markov Decision Process

SOC State of Charge

LSTM Long Short-Term Memory

RNN Recurrent Neural Network

GRU Gated Recurrent Unit

DDPG Deep Deterministic Policy Gradient

DRL Deep Reinforcement Learning

MILP Mixed Integer Linear Programming

GA Genetic Algorithms

PSO Particle Swarm Optimization

DE Differential Evolution

QoS Quality-of-Service

TOU Time-of-Use

DQN Deep Q-Network

Reinforcement Learning RL

HEV Hybrid Electric Vehicles

Acknowledgements

I wish to extend my deepest gratitude to my supervisors, Dr. Qiang Ye and Dr. Yujie Tang. Your unwavering encouragement and invaluable guidance have been instrumental in the completion of this thesis. Your continuous support and mentorship not only helped me academically but also personally. I am truly grateful for the opportunity to learn and grow under your supervision. Additionally, I would like to express my sincere appreciation to the esteemed members of the thesis committee, Dr. Peter Bodorik, Dr. Raghav Sampangi, and Dr. Meng He. The dedication of your time towards reviewing my work and your insightful feedback have been profoundly valued.

Chapter 1

Introduction

1.1 Background and Motivation

Over the past decade, the electric vehicle (EV) market has experienced an extraordinary growth trajectory. In 2021 alone, global EV sales reached a record high of 6.6 million units, resulting in a twofold increase from the previous year. This is a significant leap from the modest sales of 120,000 units in 2012 [1]. Such a robust surge in EV adoption has been primarily catalyzed by various government initiatives endorsing the transition towards cleaner, more sustainable modes of transportation [2]. Nevertheless, despite the amazing progress, concerns over extended charging periods, which often lead to range anxiety, still pose a serious barrier for many prospective EV owners [3].

In response to this challenge, Battery Swap Station (BSS) has emerged as a promising solution. These stations provide a quicker alternative to conventional charging methods, enabling fast and efficient battery replacements. Consequently, BSS has proven effective in alleviating charging anxiety among EV owners [4]. Moreover, since BSS handles battery charging and storage, it can implement an effective charging scheme to avoid peak hours and reduce charging costs, ultimately playing a significant role in peak shaving for the power grid [5].

As the reliance on BSS continues to grow, addressing the complexities associated with charging infrastructure management becomes increasingly crucial [6]. It is imperative to develop charging schemes that are both cost-effective and sustainable since they directly impact operational expenses. By dynamically adjusting the charging rates for individual batteries, BSS operators can optimize energy consumption, thereby reducing costs and improving efficiency. Considering the high power consumption of BSS, their electricity usage patterns can significantly affect the overall load profile of the grid. Substantial fluctuations in electricity demand can trigger voltage instability, accelerate the wear and tear of grid infrastructure, and in extreme scenarios, result in blackouts. Therefore, optimizing electricity usage in BSS can help maintain grid stability by smoothing the load profile and preventing grid overloading [7][4].

EV charging schemes in BSS typically fall into two categories: 1) real-time and 2) pre-determined algorithms. Pre-determined algorithms operate under the assumption that BSS operators have comprehensive knowledge of future vehicle arrival schedules. However, this ideal assumption is not valid in most practical applications, except for cases similar to electric buses that operate on fixed schedules. Conversely, real-time algorithms do not rely on prior knowledge of future information and operate conservatively, providing a worst-case performance guarantee for all potential future EV arrivals. However, several real-world studies have demonstrated that energy demands for EVs, as well as their arrival patterns, often exhibit a specific pattern. This is particularly evident when considering the phenomena of peak and off-peak hours, often associated with workplace charging [8][9]. For instance, peak hours typically occur in the early morning and late afternoon, while off-peak hours generally take place during the night. Therefore, integrating predictive information as features in a Markov Decision Process (MDP) state involved in real-time algorithms could yield substantial benefits. In a real world scenario, EVs arrive at BSS dynamically, and each EV's energy demand is addressed as much as possible. Upon arrival, the BSS operator is informed of an EV's profile. Different EVs have distinct arrival times and profiles. When determining the charging rate for each battery in a given time slot, the BSS operator needs to accurately predict the arrival times and profiles of future EVs. If we anticipate the arrival of more EVs with high energy consumption, the operator should proactively charge the available batteries at higher rates, ensuring a sufficient SOC to meet the incoming demand. Conversely, if the prediction indicates that only a few EVs with low energy demands will arrive in the near future, the operator may consider delaying the charging of some batteries. This approach combines the adaptability of real-time algorithms with valuable insights derived from predictive models, potentially leading to more effective and efficient EV charging strategies.

Devising an efficient charging scheme to manage battery charging speed is a complex problem. While existing studies primarily focus on swapping fully charged batteries with EVs, this thesis explores the possibility of providing non-fully charged batteries to EVs as long as the batteries can meet the EVs' energy requirements. With this alternative approach, BSS could offer a more personalized and efficient charging service, catering to individual EVs' energy needs and minimizing the overall energy cost.

1.2 Overview of RLC

The objective of this thesis is to develop a charging scheme to control charging speeds in BSS and minimize the cost of BSS while satisfying EV energy demand. We propose an innovative charging scheme for BSS called Reinforcement Learning-based Charging (RLC). This framework consists of two primary components: a prediction module and a charging strategy optimization mechanism.

In the prediction module, we utilize an ensemble method that combines the capabilities of Long Short-Term Memory (LSTM), Recurrent Neural Network (RNN), and Gate Recurrent Unit (GRU) models. By harnessing the strengths of these models, the ensemble learning technique provides robust and accurate predictions. LSTM, RNN, and GRU are neural network architectures specifically designed for dealing with sequential data, which makes them ideal for prediction tasks [10]. In a BSS, predicting future demands involves understanding temporal patterns, including anticipate demand and effectively schedule battery charging, which these models excel at.

For charging strategy optimization, we employ the Deep Deterministic Policy Gradient (DDPG) algorithm, a Deep Reinforcement Learning (DRL) algorithm. This algorithm is specifically designed to minimize charging costs while addressing the energy needs of EVs, generating continuous actions that correspond to specific charging speeds for each battery. Despite the challenges posed by large-scale state and action spaces and the utilization of real-time data, the DDPG algorithm offers a scalable and efficient solution for optimizing BSS charging strategies.

In the context of the Markov Decision chain, the ensemble method for prediction serves as a feature extraction mechanism. By incorporating predicted data on EV arrivals and battery swap loads into the state representation, the DDPG algorithm can generate well-informed and optimal charging strategies. This synergy between the prediction module and the reinforcement learning agent empowers the framework to adapt to fluctuations, such as variations in car arrival patterns and EV energy demand. By updating the model's understanding of the environment and effectively responding to fresh data, the framework provides a dynamic and adaptive solution for optimizing BSS charging strategies.

In this thesis, we propose a strategy to regulate charging rates within a BSS with the goal of reducing operational expenses while satisfying EV energy requests. This strategy involves the application of an RLC scheme. Initially, we formulate the relevant charging optimization problem. Following that, we utilize a predictive model to anticipate both the total EV electricity load and the number of EV batteries required. Subsequently, we extract key parameters from the EV, BSS, and grid, constructing a vector that succinctly represents the current state of the BSS while incorporating the predicted information. In the next phase, we design a DDPG network that takes the BSS state vector as input and generates an optimal charging policy. We then devise a reward function to facilitate the learning process of our model. After completing these steps, we train the DDPG network to adopt a policy that optimizes the reward function, leading to an enhancement in the overall performance of the BSS. Finally, we evaluate the efficiency of our proposed RLC scheme and compare it against existing solutions using real-world BSS data.

The main contributions of this thesis are as follows:

- We propose a novel RLC scheme to optimize charging rates in a BSS, aiming to reduce operational costs while satisfying EV energy demand. Our approach incorporates short-term and long-term predictions to forecast the number of EV batteries demanded and the total EV electricity load.

- The predictive data is integrated into a DRL framework, serving as the foundation for the decision-making process. By utilizing a DDPG network, our scheme dynamically determines the optimal charging policy. This policy is trained to maximize a carefully designed reward function, which significantly enhances BSS performance.

- In contrast to conventional methods, our scheme explores the possibility of providing EVs with non-fully charged batteries, as long as these batteries can fulfill the energy requirements of the EVs. This approach introduces a significant paradigm shift in BSS operations, offering potential cost and energy savings.

1.3 Thesis Outline

The rest of this thesis is organized as follows:

Chapter 2 provides a review of related works on BSS charging optimization, ensemble learning techniques, and DRL algorithms.

Chapter 3 introduces the DDPG algorithm for generating optimal charging strategies in BSSs. We explain the fundamentals of the algorithm and describe the architecture of the actor and critic networks. Additionally, we present the proposed ensemble method that integrates LSTM, RNN, and GRU models for predicting EV arrivals and EV swapping load at BSS. We also outline the process of incorporating predicted information into the Markov Decision Process (MDP) framework.

In Chapter 4, we present the experimental setup, describe the dataset used, and define the evaluation metrics used to assess the performance of the proposed approach. We provide a detailed analysis of the results obtained from the ensemble method for predicted information and the DDPG algorithm for charging scheme optimization. Furthermore, we compare the proposed approach with existing methods.

The thesis concludes in Chapter 5 with the Conclusion and Future Work section, where we summarize the main findings and contributions of this research. We discuss the limitations of the current work and propose potential avenues for future research to further advance the state of the art in BSS management and optimization.

Chapter 2

Related Work

2.1 BSS Charging

2.1.1 Current State and Limitations of BSS

At present, there are two primary methods for refueling electric vehicles (EVs): traditional charging and battery swapping. Traditional charging, which necessitates an EV being connected to a power source for an extended period, is well-established, with charging stations progressively expanding across the globe. However, battery swapping, a technique that rapidly exchanges an EV's depleted battery with a fully charged one, shows considerable promise for the future due to its distinct advantages. Battery swapping stations can accomplish a full ""recharge" in just a few minutes, dramatically reducing the wait time compared to traditional charging methods. This rapid turnaround is particularly crucial for commercial operations where vehicle downtime translates into financial losses [6]. Furthermore, battery swapping enables the centralization of charging, which could lead to more efficient power management and alleviate pressure on the power grid during peak demand periods [5].

Despite these advantages, battery swapping has its share of limitations. The deployment of a battery swapping station entails high capital and operational costs, which include expenses for real estate, battery inventory, and advanced swapping equipment. The rapidly evolving pace of battery technology presents another challenge. As newer, more efficient battery technologies continually emerge, battery swapping stations must shoulder the additional financial burden of perpetually updating their battery supplies to maintain efficiency and relevance [6].

Notwithstanding these challenges, certain companies recognize the substantial potential in battery swapping. Nio, a Chinese EV manufacturer, serves as a prime example [11]. The company has made significant investments in battery swapping technology, establishing over hundreds of such stations throughout China. These stations offer customers the option to swap their batteries in minutes, providing a unique advantage for their EVs. Nio is not confining its battery swapping technology to China alone. In 2022, the company ventured into the European market, indicating its intention to globally propagate this technology. Despite existing challenges, these developments underscore the promise that battery swapping holds for the future of EV refueling.

2.1.2 BSS Structure

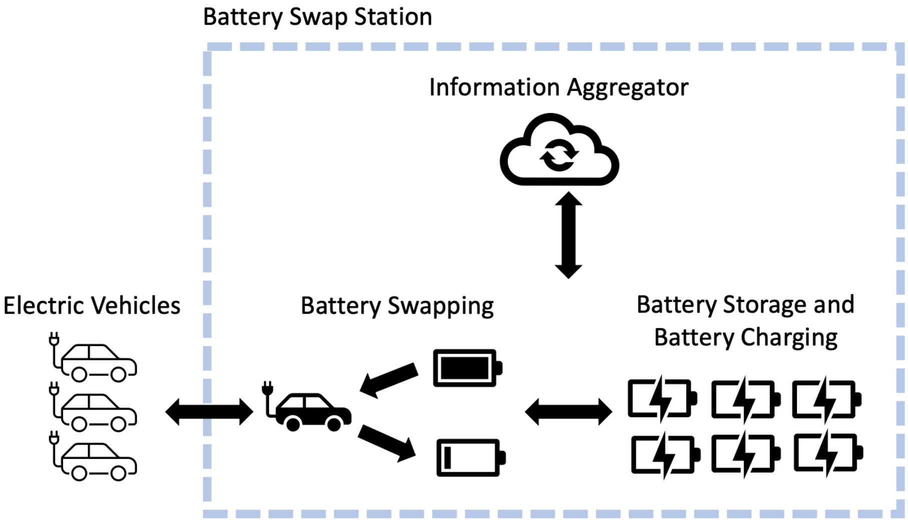

Figure 2.1: Architecture of Battery Swap Station

As depicted in Fig. 2.1, a BSS facilitates the exchange of depleted batteries in EVs for charged ones based on users' needs. The BSS comprises battery storage and charging facilities, an automated swapping mechanism, and an information aggregator. The storage and charging facilities maintain a battery inventory and multiple charging stations for recharging depleted batteries, preparing them for future swaps. The BSS features a fixed number of charging ports, each accommodating a single battery. When an EV arrives at the BSS, it requests a battery with a specific energy capacity, prompting the information aggregator to select an appropriate battery from the inventory and direct it through the swapping mechanism. The BSS information aggregator oversees all BSS operations, including battery inventory management, charging schedule coordination, and communication with the electrical grid [12].

2.1.3 BSS Charging Objectives

In recent years, there has been a growing body of research dedicated to optimizing charging strategies for BSS, with various objectives. These objectives can be broadly categorized into three main aspects: economic-related optimization, service quality-related optimization, and grid-related optimization [4]. Each of these aspects focuses on specific challenges and goals in BSS management and operation.

Economic-related optimization aims to minimize the operational costs associated with battery charging and replacement at BSSs. Approaches in this category strive to optimize battery charging and discharging schedules and determine the optimal number of batteries to minimize daily operational costs. These strategies consider factors such as electricity prices, demand charges, and battery degradation to find cost-effective charging schemes.

Service quality-related optimization focuses on enhancing the user experience by minimizing waiting times, ensuring battery availability, and meeting the energy demand of EVs. Various algorithms and techniques have been proposed to strike a balance between service quality and operational costs. These approaches take into account factors such as battery swap waiting times, battery availability, and EV energy requirements to provide efficient and reliable charging services.

Grid-related optimization centers around mitigating the impact of BSS operations on the power grid. This aspect involves optimizing charging and discharging schedules to minimize peak loads, flatten the load profile, and effectively utilize renewable energy sources. By strategically managing the charging process, BSSs can help stabilize the grid, reduce grid congestion, and maximize the utilization of renewable energy.

2.1.4 Traditional Algorithms for BSS Charging

Current research on BSS charging strategy optimization encompasses various approaches, with mixed-integer linear programming (MILP) and genetic algorithms (GA) being commonly employed. In [13], the authors focus on minimizing charging costs and energy loss while considering constraints such as bus voltage deviation, network power flow, and maximum power consumption. They propose a hybrid method that combines GA and particle swarm optimization (PSO). Similarly, in [14], the author formulates optimization objectives to maximize the BSS's battery stock level and minimize average charging damage. A comparative analysis of GA, differential evolution (DE), and PSO algorithms is conducted to solve the optimization problem.

For service quality-related optimization, in [15], the authors investigate a realistic BSS framework that addresses EV battery charging time and driving distance. Their objective is to satisfy customer demands with a quality-of-service (QoS) guarantee while considering dynamic energy pricing and varying EV arrival rates with different battery states-of-charge. They propose solutions for both online optimal BSS control and offline optimal BSS design, striking a balance between charging flexibility and battery costs. Additionally, in [16], the author utilizes a day-ahead operation method with MILP to maximize QoS scores.

While the previously mentioned works focus on QoS-related or grid-related objectives, we now turn to studies that emphasize economic-related objectives. In [17], the author aims to design a charging schema that minimizes operational costs by developing an integrated algorithm that combines the advantages of GA, DE, and PSO. The use of MILP to solve the mathematical model is a common approach. Going back to 2014, in [18], the authors develop a deterministic integer programming model to optimize the operations of battery exchange stations with the objective of minimizing operation costs. The model takes into account factors such as vehicle-to-grid technology, dependencies on power and transportation networks, and interactions between different exchange stations. An MILP approach is employed to solve the formulated optimization problem. In the same year, in [19], the authors propose an optimal scheduling model under time-of-use (TOU) electricity pricing. They utilize MILP to solve the formulated optimization problem. In 2019, in [20], the author develops a mathematical model for uncertainty-constrained BSS optimal operation. The model addresses random customer demands for fully charged batteries and leverages available batteries to reduce operation costs through demand shifting and energy sellback, while considering battery degradation for practicality.

2.1.5 DRL for BSS Charging

DRL has recently emerged as a promising approach for optimizing BSS charging schemes [21]. By combining deep learning with reinforcement learning, DRL facilitates the learning of optimal decision-making strategies in complex environments.

Unlike traditional optimization methods such as MILP and GA, DRL can adapt to dynamic changes and uncertainties in BSS operations, such as varying energy demands from different EVs and fluctuating electricity prices, making it an attractive choice for managing battery swapping and charging processes in real-world scenarios. Although DRL has been widely applied in various domains, its application in BSS charging scheme optimization is still in its early stages.

In 2019, a reinforcement learning-based charging model was developed to optimize battery swapping station operations and maximize profit [22]. This model employed a Q-learning algorithm that considered trade-offs and adapted to the varying rates of incoming vehicles. In 2020, the authors of [23] utilized DRL and proposed a BSS model that determined the optimal real-time charge/discharge power of charging piles to minimize operating costs. By implementing the DDPG algorithm, the model accounted for the stochastic operation of electric buses and the uncertainty of electricity prices, resulting in lower operating costs compared to existing benchmark control methods. In 2021, in [24], the authors modeled an individual car-sharing BSS as a coupled queuing network and implemented a Deep Q-Network (DQN) to control the charging operation of replaced batteries. This approach led to higher profits than the baseline scheme. In 2022, an Automated Guided Vehicle scenario was considered [25], involving automated guided vehicles commonly used in material handling systems with internally mounted battery packs. The proposed policy used a Markov decision process framework, and a DQN was adopted to solve the problem.

2.2 Reinforcement Learning

Reinforcement Learning (RL) is a subfield of machine learning that enables an agent to achieve its goals by interacting with its environment. In each interaction cycle, the agent selects an action based on the current state of the environment, executes the action, and receives a corresponding response from the environment, including feedback in the form of rewards and the subsequent state. This iterative process aims to maximize the expected cumulative rewards across multiple interaction cycles [26]. In certain cases where the agent's environment is fully known, the agent may not need to interact with the environment to gather data. This situation is typical in a well-defined grid world. However, this approach is not realistic for most scenarios.

In real-world reinforcement learning situations, particularly in complex physical environments, calculating state transition probabilities for actions in the Markov decision process becomes challenging. Therefore, the agent must interact with the environment and learn from the collected data, resulting in a model-free reinforcement learning approach. Value-based RL methods are a type of model-free RL that focus on learning an optimal value function. This value function estimates the expected future rewards for each state-action pair. The primary objective of these approaches is to determine the best actions to execute in each state by assessing their anticipated rewards [27]. By optimizing the value function, an agent can improve its decision-making process and enhance its overall performance within the environment.

2.2.1 Q-learning

Q-learning is a widely-used value-based RL algorithm. It updates the Q-values using the Bellman equation through iterations, combining the immediate reward with the discounted maximum Q-value for the subsequent state [28]. This iterative approach enables the agent to explore the environment randomly and gradually learn the optimal action-value function. As a result, the agent becomes well-trained and can make better decisions over time.

2.3 Deep Learning

Deep Learning, which is a distinct subset of Machine Learning, heavily relies on artificial neural networks for its operations, and often is referred to as deep neural networks. The term “deep” in Deep Learning signifies the inclusion of multiple hidden layers within the neural networks. Networks equipped with a significant number of these hidden layers are commonly referred to as “deep” [29].

In this thesis, we utilize three distinct Deep Learning models - LSTM, RNN, and GRU - to form the foundation of our ensemble learning strategy. By harnessing the unique advantages inherent in each of these Deep Learning architectures, our goal is to develop an ensemble model that not only achieves high predictive accuracy but also exhibits robustness in handling time series forecasting.

2.3.1 RNN

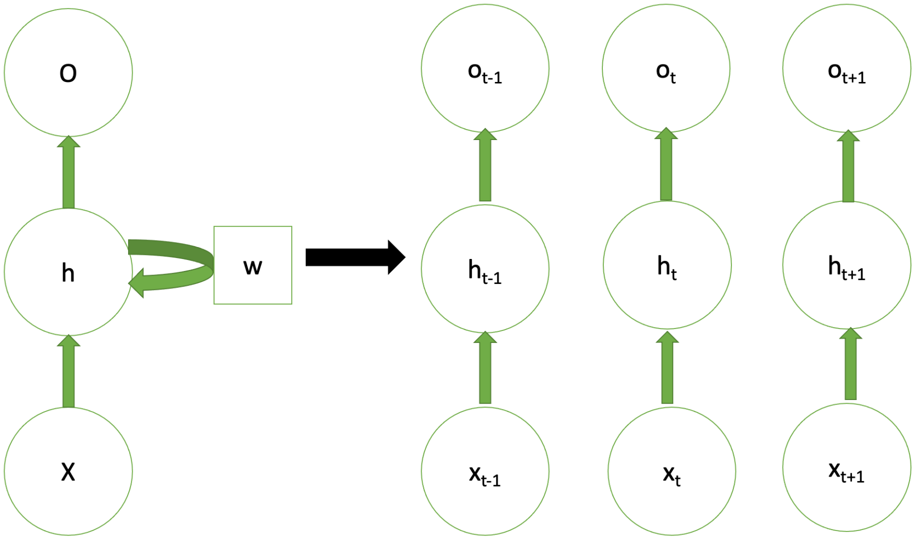

Figure 2.2: Architecture of RNN

RNN is a class of neural networks specifically designed to recognize patterns in sequential data such as text, genomes, handwriting, or time series data [30]. The architecture of RNN, represented in Fig. 2.2, is characterized by the presence of loops, enabling information to persist from one step in the sequence to the next - a feature that gives the network its "recurrent" attribute. This feature makes RNN particularly apt for tasks where the sequence of elements matters, such as language modeling or time series prediction. However, RNN have a noted limitation - their inability to handle long-term dependencies due to the "vanishing gradient" problem [31]. This issue arises during the training of an RNN using gradient-based methods when the gradient signal may become vanishingly small, resulting in slow learning or a complete cessation of learning.

2.3.2 LSTM

LSTM networks, illustrated in Fig. 2.3, are a subtype of RNNs, designed specifically to address the vanishing gradient problem [32]. They manage this through the implementation of a sophisticated cell state capable of retaining information in memory over extended periods. This feature makes LSTM suitable for tasks involving sequences with long-term dependencies. Each LSTM cell consists of three gates: an

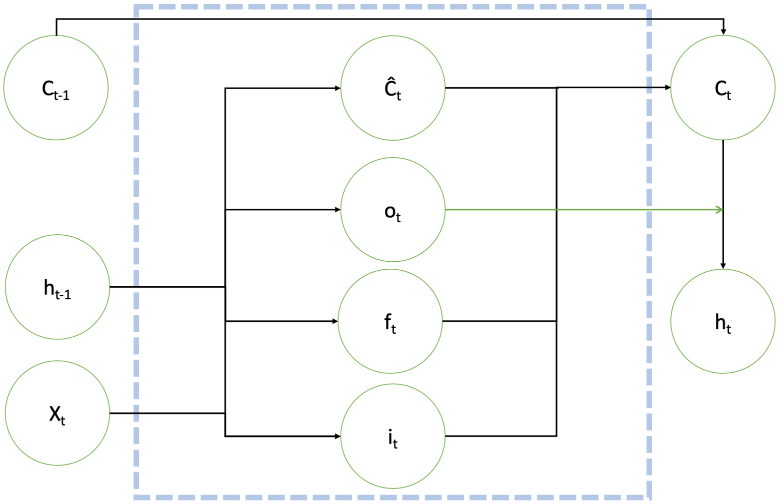

Figure 2.3: Architecture of LSTM

input gate, a forget gate, and an output gate. These gates regulate the addition of new information to the cell state, the discarding of old information, and the determination of the current output based on the cell state. This intricate gating mechanism enables LSTM to learn longer sequences and maintain an extended memory, which is an attribute not found in vanilla RNN.

2.3.3 GRU

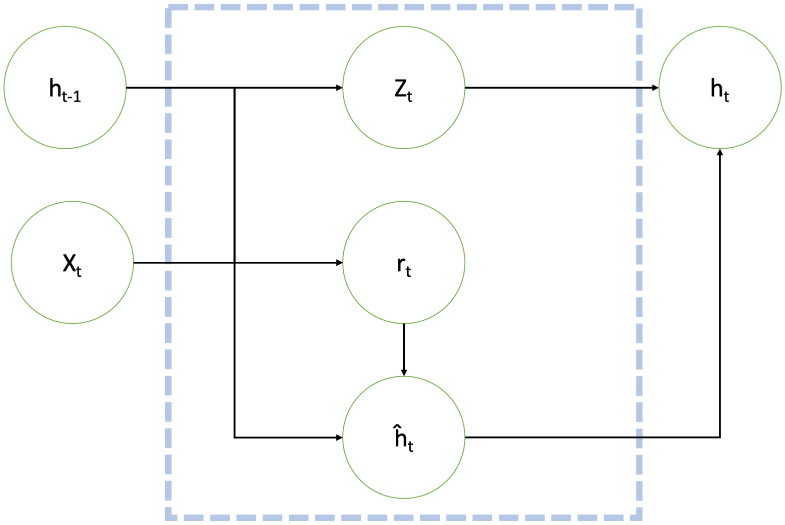

Figure 2.4: Architecture of GRU

GRU, depicted in Fig. 2.4, is designed as a simpler and more computationally efficient alternative to LSTM [33]. Similar to LSTM, GRU is also a variant of RNN that effectively addresses the vanishing gradient problem and is proficient at handling longer sequences.

A GRU cell consists of two gates: a reset gate and an update gate. The reset gate determines how to combine the new input with the previous memory, while the update gate defines how much of the previous memory to retain. While GRU has fewer parameters and thus are quicker and easier to compute, they may not perform as effectively as LSTM on tasks that require more complex memory manipulation. However, the choice between LSTM and GRU can vary depending on the specific requirements and constraints of a given task.

2.4 Deep Reinforcement Learning

In traditional RL, state-action pairs are typically stored in a mapping structure, where each action corresponds to the reward that can be obtained in the current state. However, this approach is only suitable for environments and actions with limited dimensions and discrete values. As the dimensions of the environment and actions increase, the number of possible states grows exponentially. Furthermore, in cases where the environment and actions are continuous, the number of states can become virtually infinite. This poses a significant challenge for limited computational resources to handle such scenarios.

To address this challenge, DRL has been developed. DRL algorithms leverage the power of deep neural networks as function approximators for value functions or policies. By using deep neural networks, DRL can effectively capture and represent the complex relationships between states, actions, and rewards. This capability allows DRL to be applied to a wider range of problems, including those involving high-dimensional and continuous environments and actions. In contrast to traditional RL methods, which would struggle with such cases, DRL algorithms can handle these complex and large-scale problems more efficiently [34].

2.4.1 Value and Policy Based DRL

Value-based deep reinforcement learning methods are built on the foundation of DRL and primarily focus on estimating the value function, which represents the expected cumulative reward for taking an action in a given state. By utilizing deep neural networks to approximate the value function, these methods can effectively handle high-dimensional and continuous state and action spaces.

One well-known value-based DRL algorithm is the DQN, an extension of the Q-Learning algorithm. DQN employs a deep neural network as a function approximator for the Q-function, which represents the expected cumulative reward when taking an action in a state and following the optimal policy. The goal of the DQN algorithm is to minimize the loss function, defined as the mean squared error between the predicted Q-value and the target Q-value. Through iterative updates of the neural network's parameters to minimize this loss function, DQN learns to approximate the optimal Q-function, enabling it to discover effective policies for problem-solving. To address instability and divergence issues in deep reinforcement learning, DQN introduces essential techniques such as Experience Replay and Target Network [35]. Experience Replay involves storing the agent's experiences in a replay buffer and randomly sampling mini-batches of experiences from the buffer to update the neural network during training. This method helps break correlations between consecutive samples, improving the stability of the learning process. The Target Network is a separate neural network with the same architecture as the original DQN network. It is used to compute the target Q-values during training, and its parameters are periodically updated with those of the original network. This technique mitigates the issue of moving target Q-values, which can lead to an unstable and divergent learning process.

On the other hand, policy-based deep reinforcement learning methods directly learn the optimal policy, mapping states to actions, instead of approximating the value function like value-based methods [36]. In policy-based DRL, a deep neural network typically parameterizes the policy. The policy network takes the current state as input and generates probabilities for selecting each action or the mean and standard deviation of a continuous action distribution. The objective is to optimize the policy network's parameters to maximize the expected cumulative reward.

2.4.2 From Actor-Critic Based DRL to Deep Deterministic Policy Gradient

The Actor-Critic approach is a fusion of policy-based and value-based methods within the realm of DRL, combining two components: the actor and the critic [36]. The actor interacts with the environment and selects actions based on the current policy. The policy is trained using the policy gradient method, guided by the value function provided by the critic. On the other hand, the critic approximates the value function and evaluates the actor's action choices. By providing feedback to the actor through its interaction with the environment, the critic helps fine-tune the policy to maximize the expected cumulative reward. The Actor-Critic system leverages the strengths of both policy-based and value-based methods, effectively handling the complexities of learning optimal policies in complex RL problems. It combines the stability of direct policy learning from policy-based methods with the value function estimation from value-based methods for policy updates.

DDPG algorithm [37] enhances the conventional actor-critic approach used in reinforcement learning. Traditional policy-based methods typically result in a stochastic policy and are utilized in online learning scenarios, leading to low sample efficiency. Contrarily, DQN focuses on directly estimating Q-values, excels in discrete action spaces, and is implemented as an offline algorithm. Nevertheless, DQN encounters difficulties dealing with continuous action spaces. DDPG was conceived to address these specific issues concurrently found in both policy-based and value-based methods. DDPG employs a deterministic policy, contrary to the traditional stochastic ones, and is designed to optimize the Q-value using a gradient ascent process. This amalgamation makes DDPG a powerful tool for dealing with continuous action spaces and promoting sample efficiency.

2.4.3 Integrating Predictive Information into DRL

Existing studies, such as [38] and [39], have demonstrated the benefits of integrating predictive information into DRL to enhance performance and robustness of reinforcement learning algorithms. By incorporating predictive information directly into the input features, DRL algorithms can leverage this information during the decision-making process. This approach has been widely adopted in various domains, including finance and energy sectors.

In [38], the authors propose a predictive energy management strategy for parallel hybrid electric vehicles (HEVs) using velocity prediction and RL. The approach involves modeling the HEV, defining a cost function, and employing fuzzy encoding and nearest neighbor techniques for velocity prediction. The strategy also utilizes a finite-state Markov chain to learn power demand transition probabilities and determine optimal control behaviors and power distribution between energy sources. The look-ahead energy management strategy is compared to shortsighted and dynamic programming-based methods, and the results show that the RL-optimized control effectively reduces fuel consumption and computational time. In [39], the authors present a deep RL architecture for automating dynamic portfolio optimization. The proposed model incorporates an infused prediction module, a generative adversarial data augmentation module, and a behavior cloning module. It works with both on-policy and off-policy RL algorithms and interacts with a back-testing and execution engine in real time. The infused prediction module helps capture predictive information, enhancing the model's ability to make informed decisions for portfolio optimization.

2.5 Ensemble Learning

Time series forecasting has extensive applications in real-world scenarios, including energy load prediction [40][41] and EV arrival pattern prediction [42]. As machine learning continues to advance, deep learning, which is a subset of machine learning, is gaining increasing popularity in the field of time series forecasting. Existing research highlights the effectiveness of specific deep learning algorithms for time series forecasting, such as RNN [43], LSTM networks [44], and GRU [45]. These models have demonstrated their ability to identify complex patterns and trends in time series data.

Ensemble learning is an approach that improves the overall accuracy of deep learning models by integrating the strengths of individual models and mitigating their weaknesses [46]. This method combines the predictions of multiple base models to generate robust and accurate predictions. The underlying principle of ensemble learning is that a diverse group of models, each with unique strengths, can collectively outperform any individual model. Ensemble learning techniques, such as averaging and weighted averaging, are used to combine the predictions from the base models [47].

This chapter presents the background of BSS charging. Additionally, we select three foundational models - LSTM, RNN, and GRU - for our ensemble learning strategy. The decision to employ this strategy, rather than depending on a single model, is influenced by the principle of diversity in model building, a concept explored by Kuncheva [48]. This method acknowledges each model's unique strengths and weaknesses, understanding that their integration can result in a prediction system that is both comprehensive and robust. Moreover, ensemble learning adds a layer of robustness. By aggregating the predictions of multiple models, we can mitigate the effect of shortcomings from any individual model on the overall prediction [46]. Thus, if one model inaccurately predicts a particular trend, others in the ensemble may compensate for it, leading to a collectively more precise prediction. For the RL agent, we choose DDPG due to its successful merging of both actor and critic network advantages. In the following chapter, we will delve into the specifics of our proposed BSS charging schemes.

Chapter 3

Reinforcement Learning Based Battery Charging

3.1 System Model and Problem Formulation

In this section, we first outline the underlying assumptions of our study. Following this, we present the problem formulation for our BSS optimization research. Key notations used throughout this paper are summarized in Table 3.1. Detailed explanations of these notations will be provided in the subsequent sections of this paper.

Table 3.1: Key Notations

Notation	Description
S	State set
s	State
t	Time slot
d_t	Number of demanded batteries
e_t	Battery swapping load
$\hat{d}_t$	Predicted number of demanded batteries
$\hat{\boldsymbol{e}}_t^{total}$	Predicted total battery swapping load
$SOC_{i,t}^{BSS}$	Charging levels of i_{th} batteries in BSS
$SOC_{j,t}^{EV}$	Charging levels of j_{th} incoming EVs
a_t	Charging rates

3.1.1 Assumptions

To simplify the problem and facilitate modeling, we make the following assumptions:

1. All EVs are compatible with the BSS, and their batteries can be swapped using the available mechanism.

2. The time required for the battery swapping process is negligible compared to the time spent on charging the batteries.

3. We disregard the degradation of the batteries over time.

4. If no available batteries in the BSS meet or exceed the energy requests of the EVs, we assume that each EV has a predefined threshold level. This threshold level is larger than their current battery level, and a battery swap will only occur if an available battery exceeds this threshold.

5. All batteries are assumed to have identical properties - they possess the same maximum charge capacity, undergo the same rate of degradation, and experience the same loss of power or charge for a given amount of charge consumed.

6. The charging characteristics are assumed to be identical for each battery. This implies that all batteries take the same amount of time to charge from a given SOC to another, assuming the same charging power is used.

3.1.2 Problem Formulation

In this section, we present the problem formulation for the BSS charging optimization problem under investigation. The key elements of the problem are as follows:

BSS: The BSS is equipped with N batteries and M charging ports. We assume that the number of batteries equals the number of charging ports, represented as $M = N$.

Time-series State: The system operates over a series of time slots represented by T, with each time slot defined as $t \in T$. At each time slot t, the state s_t encompasses both internal BSS information, such as the State of Charge (SOC) of stock batteries and external information from arriving EVs, including arrival time, SOC level, and grid electricity price. The collection of states is denoted as S. The state s_t can be represented as a Markov chain, where $s_t \in S$ and S is the set of all possible states.

EV Arrival: At the onset of each time slot t, d_t EVs arrive at the BSS. The state s_t is derived based on the number of demanded batteries d_t and the battery swapping load of the EVs e_t, with $e_t = [e_{1,t}, e_{2,t}, \ldots, e_{d_t,t}]$.

SOC: The charging level of the i^{th} battery in the BSS at time t is represented as $SOC_{i,t}^{BSS}$, while the initial charging level of the j^{th} EV, where j ranges from 0 to d_t, arriving at the BSS during time slot t is denoted as $SOC_{j,t}^{EV}$.

Battery Swapping Process: The battery swapping process is outlined in Alg. 1. At the end of each time slot t, batteries within the BSS that is closest to, but not less than the user's energy request e_t are swapped for the j-th EV. If no batteries fulfill this condition, a battery in the BSS with a SOC larger than the sum of the current EV's SOC and a defined threshold θ is swapped. This occurs when $SOC_{i,t}^{BSS}$ is greater than $SOC_{j,t}^{EV} + \theta$. If neither of these conditions are met, the EV leaves the BSS without a battery swap.

Algorithm 1 Battery Swapping Process

```
Require: Battery demand $d_t$ and EV requests $e_t = [e_{1,t}, e_{2,t}, \dots, e_{d_t,t}]$
Let $SOC_{i,t}^{BSS}$ denote the SOC of the $i^{th}$ battery in the BSS, and $SOC_{j,t}^{EV}$ denote the SOC of the $j^{th}$ EV;
for $j = 1$ to $d$ do                                    ▷ Loop over all EV requests
    Find a battery whose $SOC^{BSS} i, t$ is closest to, but not less than $ej, t$;
    if no such battery is found then
        Find a battery whose $SOC_{i,t}^{BSS}$ is closest to, but not less than $SOC_{j,t}^{EV} + \theta$;
    end if
    if no such battery is found then
        The $j^{th}$ EV leaves without a battery swap;
    end if
    Update $SOC_{i,t}^{BSS}$ if there was a battery swap;
end for
```

Battery Charging: The charging rate of the BSS at time t, denoted as $a_{i,t}$, can be adjusted by an RL agent. The charging constraints are formulated as follows:

$$SOC_{\min} \leq SOC_{i,t}^{BSS} \leq SOC_{\max} \quad \forall i, t \tag{3.1}$$

$$P_{\min} \leq a_{i,t} \leq P_{\max} \quad \forall i, t \tag{3.2}$$

where $SOC_{\min}$ and $SOC_{\max}$ are the minimum and maximum battery SOC limits, and $P_{\min}$ and $P_{\max}$ are the minimum and maximum charging rates of the BSS. The constraint (3.1) ensures that the battery SOC stays within acceptable levels, while constraint (3.2) ensures that the charging rates of the BSS stay within acceptable limits.

Objective: The main objectives of the RL agent are:

- Minimize Electricity Cost: The RL agent should adjust the charging powers of the batteries to minimize overall electricity cost. This involves taking advantage of fluctuating electricity prices to charge more during off-peak times and less during peak times.

- Fulfill EV Battery Replacement Demand: The RL agent should aim to meet the energy requests of arriving EVs as much as possible. This requires making sure there are enough batteries with the necessary SOC available when EVs arrive.
- Maintain High SOC for BSS Batteries: The RL agent should also aim to keep the batteries in the BSS as fully charged as possible. This provides a buffer to meet unexpected surges in demand. However, this must be balanced against the objective of minimizing electricity cost, since keeping batteries fully charged could involve charging during peak times.

3.2 Details of RLC

In our research, we explore the application of DRL for optimizing the BSS charging scheme with the goal of minimizing costs while meeting the energy requests of EVs. While existing research has explored the use of DRL for BSS optimization, this thesis represents the first attempt to incorporate predicted results and DRL to tailor charging strategies based on needed SOC from users. We employ the DDPG algorithm to learn optimal charging policies by observing system states from the environment, EV battery swapping loads, EV battery demand numbers from the prediction model, and their impact on battery inventory and charging facilities.

The proposed RLC scheme's workflow, depicted in Fig. 3.1, integrates the environment, BSS information representation, a prediction module, and an RL agent. The prediction module predicts certain information of following time slots, such as total EV electricity load and EV battery demand numbers. To provide further clarification, we have defined two terms: when we predict the immediate next time slot, we term it as RLC with Short Term Prediction (RLC-S), and when we are predicting multiple future time slots, we refer to it as RLC with Long Term Prediction (RLC-L). Detailed explanations of Short Term Prediction (STP) and Long Term Prediction (LTP) will be provided in the subsequent sections.

In addition, the RL agent observes the environment to make informed decisions. Based on these inputs, the RL agent generates actions for continuous charging rates

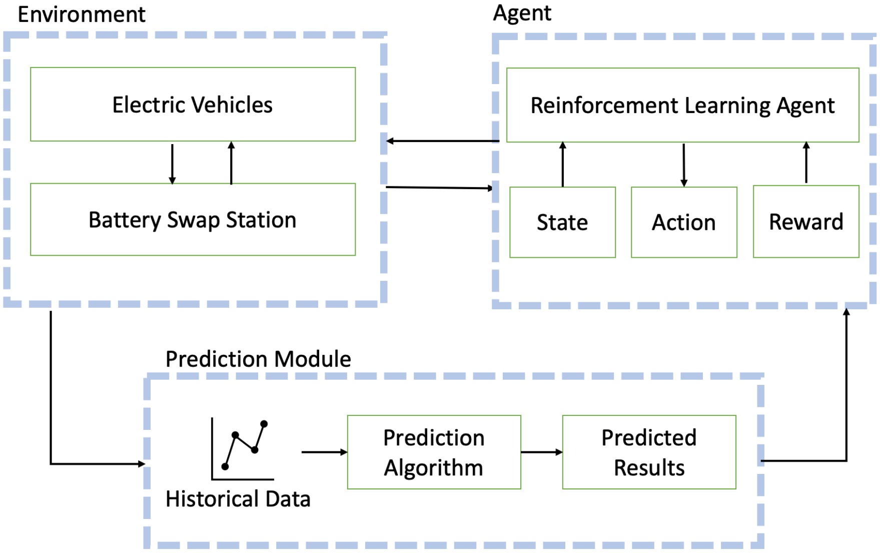

Figure 3.1: Workflow of RLC

and receives rewards from the environment. These policies aim to prioritize battery charging processes, taking into account available resources in BSS. For instance, during peak usage of the BSS, the RLC scheme prioritizes efficiently charging batteries to a SOC level that corresponds with the expected energy demands of incoming EVs. This strategy ensures a smooth and uninterrupted battery swapping experience. Conversely, when the BSS has ample available resources, the scheme adapts its charging policies to accommodate a broader range of energy requests, including those from vehicles with high energy demands.

The rest of this section delves into the details of the RLC scheme.

3.2.1 Prediction Module

Incorporating predictive information into the agent's state can offer substantial benefits. It enables the agent to focus more effectively on observations and experiences, reducing the need to interpret or anticipate system uncertainties. To incorporate this predictive information, we employ a prediction module. While the RL agent doesn't inherently require prior knowledge or assumptions regarding system uncertainties, the integration of a predictive model can provide the agent with a form of "anticipated" state. This anticipated state can be invaluable in the agent's decision-making process. The objective of this predictive information is to present a probable future system state. It serves as a guidepost for the RL agent's learning process, effectively offering the agent a glimpse into a possible future.

Short-term Prediction vs Long-term Prediction

The prediction module is specifically designed to forecast the total electricity load and the number of battery demand, which assists the DRL agent in making more informed decisions. The total electricity load represents the aggregate energy demanded by EVs at each time slot, while the number of demanded battery requests indicates the quantity of EVs arriving at the BSS and requesting a battery swap. Given a time series of data points $T_{n-t_s}, T_{n-2}, T_{n-1}, \ldots, T_n$, where t_s represents the time step of historical data, our prediction module aims to predict the subsequent time slots of total electricity load and number of demanded battery for both short-term and long-term predictions.

For STP, we predict the immediate next time slot T_{n+1} by analyzing a length of historical data t_s. This enables the DRL agent to anticipate and make decisions based on the upcoming time slot. For LTP, we predict l numbers of future time slots $T_{n+1}, T_{n+2}, \ldots, T_{n+l}$ by analyzing the same length of historical data t_s. This allows the DRL agent to anticipate and plan for trends and changes in battery swapping demand over a longer time horizon.

Details of Prediction Module

The workflow of our proposed ensemble learning-based forecasting module is depicted in Fig. 3.2. Ensemble learning is employed in this research due to its ability to improve prediction accuracy and generalization by combining the strengths of multiple base models [49]. The initial step involves training the three base models, RNN, LSTM, and GRU. After training the base models, their predictions are generated. To find appropriate weights for our base models, Linear Regression (LR) is adopted for the weighted average of the base models. The purpose of the LR model is to learn the optimal weights for each base model, balancing their contributions to the final predictions [50]. The equation for LR can be expressed as:

$$y = w_1 \cdot y_1 + w_2 \cdot y_2 + \cdots + w_n \cdot y_n + \epsilon \tag{3.3}$$

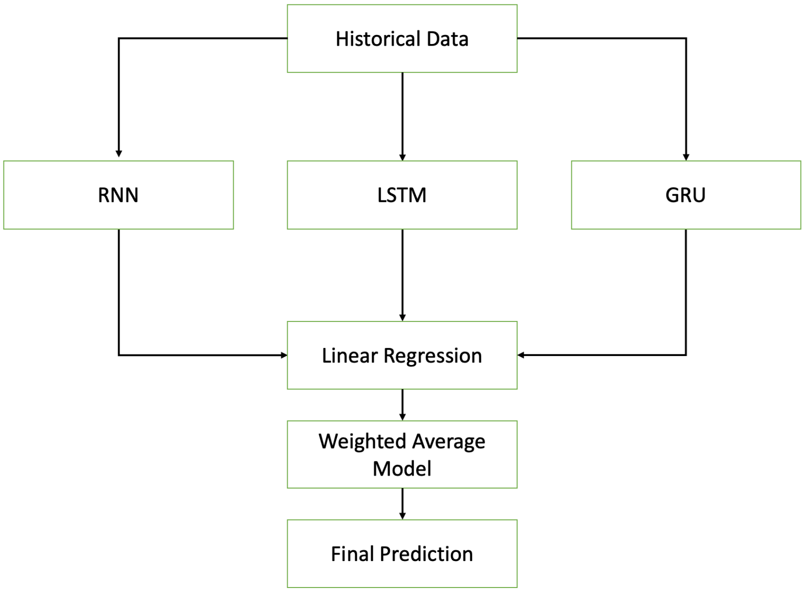

Figure 3.2: Workflow of Prediction Module

Where y is the target variable, y_i are the meta-features (in this case, predictions from the base models), and w_i are the learned weights. The predictions and true labels of the training data are set as the meta-features and target for LR training, respectively. By minimizing the discrepancy ϵ between the predicted output and the actual output, LR effectively learns the optimal weights for each base model (RNN, LSTM, GRU).

Once the LR model has learned the appropriate weights for each base model (RNN, LSTM, GRU), a weighted average model is employed to combine the predictions of the base models. The equation for the weighted average model is as follows:

$$y_{final} = w_1 \cdot y_1 + w_2 \cdot y_2 + \cdots + w_n \cdot y_n \tag{3.4}$$

Where y_{final} is the final prediction obtained from the weighted average model, y_i are the predictions from the base models, and w_i are the learned weights from the LR model.

Alg. 2 provides a detailed process of our forecasting model. The process begins by training the base models (RNN, LSTM, and GRU) on the training data, generating predictions for each instance in the training data using these models. These predictions, which serve as meta-features, are then used to train the LR model. The LR model learns the optimal weights for each base model by minimizing the discrepancy between the weighted average of the base models' predictions and the actual labels

Algorithm 2 Ensemble Learning Algorithm for Total Electricity Load and Number of Demanded Batteries

Require: Training data $D_{Train} = (X_1, Y_1), (X_2, Y_2), \ldots$, test data $D_{Test} = (x_1, y_1), (x_2, y_2), \ldots$, base models M_i, where $i = 1, 2, 3$, Linear Regression model LR, Ensemble learning model M_{LR}

```
for $X_i$ in $D_{Train}$ do
    for $i = 1, 2, 3$ do
        $P_{i,Train} \leftarrow M_i(X_i)$
    end for
end for
$P_{Train} \leftarrow [[P_{1,Train}], [P_{2,Train}], [P_{3,Train}]]$
$M_{LR} \leftarrow LR(P_{Train}, Y)$
for $x_i$ in $D_{Test}$ do
    for $i = 1, 2, 3$ do
        $P_{i,Test} \leftarrow M_i(x_i)$
    end for
end for
$P_{Test} \leftarrow [[P_{1,Test}], [P_{2,Test}], [P_{3,Test}]]$
$P_{Final} \leftarrow M_{LR}(P_{Test})$
```

in the training data. Once the LR model is trained, it is used to compute the final predictions for the testing data. Specifically, each base model makes a prediction for each instance in the testing data, and these predictions are input into the LR model. The LR model then generates the final prediction for each instance by taking a weighted average of the base models' predictions, with the weights determined by the coefficients it learned during training. The accuracy of these final predictions is then evaluated against the true labels of the testing data.

3.2.2 MDP for Charging Optimization

The optimization problem can be formulated as an MDP, defined by the tuple $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R})$.

State

The state space ($\mathcal{S}$) represents the relevant information required for decision-making in the BSS. A state s_t at time slot t is given by the following components:

$$s_t = (p_t, d_t, SOC_{i,t}^{BSS}, SOC_{j,t}^{EV}, \boldsymbol{e}_t) \tag{3.5}$$

In the state space, p_t represents the electricity price at time slot t, d_t denotes the number of demanded batteries, $\mathrm{SOC}_{i,t}^{\mathrm{BSS}}$ refers to the battery SOC levels stored in the BSS at time slot t, $\mathrm{SOC}_{j,t}^{\mathrm{EV}}$ stands for the initial EV SOC (a vector representing the charging levels of batteries for arrived EVs at time slot t), and $\boldsymbol{e}_t$ signifies the battery swapping load at time t (a vector representing the energy that each EV requests at time slot t).

When incorporating both STP and LTP into the states, we can update the states to include the predicted battery demand and the predicted aggregated EV swapping load for the next time slot $t+1$ for STP, and for the next n time slots for LTP. The DRL agent has the ability to access both short-term and long-term forecasts. This capability enables it to make decisions that are not only informed by the immediate future but also by potential trends and patterns in the battery swapping load and the number of battery swap requests over an extended time frame. The states can then be revised as follows:

$$s_t^{STP} = (p_t, d_t, SOC_{i,t}^{BSS}, SOC_{j,t}^{EV}, \boldsymbol{e}_t, p_{t+1}, SOC_{ave}^{EV}, \hat{d}_{t+1}, \hat{\boldsymbol{e}}_{t+1}^{ave}) \tag{3.6}$$

$$s_t^{LTP} = (p_t, d_t, SOC_{i,t}^{BSS}, SOC_{j,t}^{EV}, \boldsymbol{e}_t, p_n, SOC_{ave}^{EV}, \hat{d}_n, \hat{\boldsymbol{e}}_n^{ave}) \tag{3.7}$$

$$\hat{\boldsymbol{e}}_t^{ave} = \frac{\hat{\boldsymbol{e}}_t^{total}}{\hat{d}_t} \tag{3.8}$$

Within these updated state expressions, represented by Eq. (3.6) for STP and Eq. (3.7) for LTP, $SOC_{ave,t}^{EV}$ denotes the average SOC upon arrival at the BSS. Meanwhile, $\hat{\boldsymbol{e}}_t^{ave}$ signifies the projected average energy request from the EV, which is calculated based on the predicted number of required batteries and the predicted total electricity load as described in Eq. (3.8).

In the case of STP, p_{t+1} is the electricity price in the next time slot. $\hat{d}_{t+1}$ indicates the expected number of batteries required, providing a prediction of battery demand for the next time slot. $\hat{\boldsymbol{e}}_{t+1}^{ave}$ represents the projected average energy request from the EV for the next time slot.

In contrast, for LTP, p_n is a vector contains electricity price for the next n time slots. $\hat{d}_n$ is a vector that embodies the predicted information about the number of batteries needed for the subsequent n time slots. Similarly, $\hat{\boldsymbol{e}}_n^{ave}$ is a vector that captures the projected average energy requested by the EV for the next n time slots.

Action

The action space ($\mathcal{A}$) is composed of continuous charging rates for each charging port at BSS. Let m represent the number of available charging ports in the BSS. Consequently, the action a_t at any given time slot t constitutes a vector of charging rates across all these ports:

$$a_t = (a_{1,t}, a_{2,t}, \ldots, a_{m,t}) \tag{3.9}$$

Each charging rates $a_{m,t}$ in the vector is constrained by the maximum charging rate P_{max}:

$$P_{min} \leq a_{m,t} \leq P_{max} \tag{3.10}$$

State Transition

The state transition ($\mathcal{P}(s_{t+1}|s_t, a_t)$) is used to describe the state transition process in the BSS system. The algorithm for the state transition process is shown in Alg. 3. The algorithm takes the current state s_t and the action a_t as input and produces the next state s_{t+1} as output. The state transition process includes observing the electricity price p_{t+1}, battery demand d_{t+1}, initial EV SOC $\mathrm{SOC}_{j,t+1}^{\mathrm{EV}}$, and EV request energy $\boldsymbol{e}_{t+1}$. Meanwhile, it updates the SOC of the BSS $SOC_{i,t+1}^{\mathrm{BSS}}$ based on swapped batteries from EV swapping load $\boldsymbol{e}_t$, current $SOC_{i,t}^{\mathrm{BSS}}$, and charging rates a_t. The swapping scheme Alg. 1 was introduced in the previous section. Next, the predicted aggregated EV swapping load $\hat{\boldsymbol{e}}_{t+1}$ and predicted battery demand $\hat{d}_{t+1}$ are obtained using the predicted model M_p. Finally, the state s_{t+1} is set to the current state s_t with the updated values of p_{t+1}, d_{t+1}, $SOC_{i,t+1}^{\mathrm{BSS}}$, $SOC_{j,t+1}^{\mathrm{EV}}$, e_{t+1}, $\hat{d}_{t+2}$, and $\hat{\boldsymbol{e}}_{t+2}$.

Algorithm 3 State Transition

```
Require: $s_t$, $a_t$
Ensure: $s_{t+1}$
1: $d_{t+1} \leftarrow d_t$
2: $SOC^{EV}_{i,t+1} \leftarrow SOC^{EV}_{i,t}$
3: $p_{t+1} \leftarrow p_t$
4: $e_{t+1} \leftarrow e_t$
5: $SOC^{BSS}_{i,t+1} \leftarrow [SOC^{BSS}_{i,t}, SOC^{EV}_{i,t}, e_t, a_t]$
6: $(\hat{d}_{t+2}, \hat{\boldsymbol{r}}_{t+2}) \leftarrow M_p$
7: $s_{t+1} \leftarrow s_t$
```

Reward Function

The reward function ($\mathcal{R}$) calculates the immediate benefit or cost associated with taking action a_t in state s_t. In this case, the reward function is designed to minimize three key components: the electricity cost, the discrepancy between the desired and provided battery levels, the discrepancy between the battery full capacity and the current capacity. In DDPG, the agent aims to minimize the negative sum of these three components, which can be formulated as follows:

$$\mathcal{R}(s_t, a_t) = -(C_{elec} + C_{demand} + C_{capacity}) \tag{3.11}$$

C_{elec} represents the electricity cost, calculated as the difference between the total electricity cost at current price p_t and the total cost at the minimum price p_{min}, considering the charging rate a_t for each charging port m, where $m \in 1, 2, ..., M$, over the duration of the time slot Δt, the charging efficiency η and penalty weight α:

$$C_{elec} = \alpha \cdot max(0, (\sum_{m=1}^{M} \frac{a_{m,t}}{\eta} \cdot p_t \cdot \Delta t - \sum_{m=1}^{M} \frac{a_{m,t}}{\eta} \cdot p_{min} \cdot \Delta t)) \tag{3.12}$$

The goal of Eq. (3.12) is to minimize the cost of electricity used for charging the BSS. In an environment where electricity prices fluctuate over time, charging strategies can significantly impact the overall cost. Charging more when electricity prices are low and less when prices are high can reduce costs. The equation for C_{elec} is designed to encourage the model to optimize the charging speed ($a_{m,t}$) in response to price fluctuations. If the cost at the current electricity price (p_t) is higher than the cost at the minimum electricity price ($p_{\min}$), the model is encouraged to reduce the charging speed to decrease the electricity cost. On the other hand, if the current price is less than or equal to the minimum price, the model is encouraged to charge at a higher speed to take advantage of the lower price. By evaluating the additional cost incurred due to price fluctuations, this component of the reward function encourages the adoption of a cost-effective charging strategy.

Indeed, the term C_{demand} in Eq. (3.13):

$$C_{demand} = \beta \cdot \sum_{j=0}^{d_t} \max(0, SOC_{j,t}^{Desire} - SOC_{j,t+1}^{BSS}) \cdot B \tag{3.13}$$

$$SOC_{j,t}^{Desire} = \frac{SOC_{j,t}^{EV} \cdot B + \boldsymbol{e_t}}{B} \cdot 100\% \tag{3.14}$$

represents the potential additional cost that the BSS could incur when it fails to meet the energy demands of the arriving EVs in the current time slot. In Eq. (3.14), $SOC_{j,t}^{Desire}$ is the desired state of charge for the j^{th} EV, calculated by adding the initial energy amount of the EV to the requested energy amount and then converting it to SOC. $SOC_{j,t+1}^{BSS}$ represents the state of charge of the j^{th} battery in the BSS, and B denotes the battery capacity. The difference between the desired and available SOC, given by $\max(0, SOC_{j,t+1}^{Desire} - SOC_{j,t+1}^{BSS})$, and multiplied by the battery capacity, provides the additional energy required to achieve the desired state of charge. The factor β is a penalty weight that can be adjusted to control the emphasis placed on meeting the EVs' energy demand in the overall cost function. This design equips the RL agent with the capability to take into account the cost of unmet demand. By doing so, it can adapt its charging strategy to strive for fulfilling the energy requests of the EV.

The third component, $C_{capacity}$, defined in Eq. (3.15):

$$C_{capacity} = \gamma \cdot \sum_{m=1}^{M} \max(0, SOC_{max} - SOC_{m,t+1}^{BSS}) \cdot B \tag{3.15}$$

represents the potential cost associated with the discrepancy between the maximum battery capacity and its current state. Here, γ is a penalty weight that can be adjusted to control the emphasis placed on maintaining the battery's state of charge close to its maximum capacity in the cost function. This component of the reward function encourages the RL agent to maintain an appropriate level of battery state of charge, even in periods when no vehicles are arriving at the BSS. It's a mechanism designed to ensure readiness to meet future demand, thereby enhancing the overall service efficiency and battery lifespan.

Reward Function with STP

To incorporate STP, we add two additional components. The first component is the potential additional electricity cost for STP, denoted as C_{elec}^{STP}:

$$C_{elec}^{STP} = \alpha^{'} \cdot \max(0, \sum_{m=1}^{M} \frac{a_{m,t}}{\eta} \cdot p_t \cdot \Delta t - \sum_{m=1}^{M} \frac{a_{m,t}}{\eta} \cdot p_{t+1} \cdot \Delta t) \tag{3.16}$$

In Eq. (3.16), C_{elec}^{STP} represents the potential additional electricity cost for STP, which is calculated based on the difference between the total electricity cost at the current price p_t and the total cost at the predicted price for the next time slot p_{t+1}, considering the charging rate a_t for each charging port m over the duration of the time slot Δt and the charging efficiency η. $\alpha^{'}$ is a penalty weight adjusted for STP. This term encourages the model to adjust the charging speed ($a_{m,t}$) in response to predicted future electricity price fluctuations. If the current cost is higher than the predicted cost for the next time slot, the model is encouraged to reduce the charging speed to decrease the electricity cost, and vice versa.

The second component is the potential supplementary cost that the BSS might incur if it fails to meet the anticipated energy needs of the forthcoming EVs in the next time slot, denoted as C_{demand}^{STP}:

$$C_{demand}^{STP} = \beta^{'} \cdot \sum_{i=0}^{\hat{d}_{t+1}} \max(0, SO\hat{C}_t^{Desire} \cdot B - (SOC_{i,t+1}^{top} + \Delta SOC_{max}) \cdot B) \tag{3.17}$$

$$SO\hat{C}_{t+1}^{Desire} = SOC_{ave}^{EV} + \frac{\hat{\boldsymbol{e}}_{t+1}^{ave}}{B} \cdot 100\% \tag{3.18}$$

$$SOC_{i,t+1}^{top} = SOC_{t+1}^{BSS,[i]} \quad \text{for} \quad i \in 0, 1, 2, ..., \hat{d}_{t+1} \tag{3.19}$$

In Eq. (3.18), the predicted average desired electricity SOC for the EV battery is obtained by adding the average SOC of EV battery upon arrival at the BSS, and the projected average energy demand SOC from the EVs. In Eq. (3.17), $\hat{d}_{t+1}$ represents the predicted battery demand numbers. ΔSOC_{max} is the maximum permissible change in SOC within a single time slot. The top $\hat{d}_{t+1}$ SOC of the BSS for the next time slot is defined as $SOC_{i,t+1}^{top}$ in Eq. (3.19). By accumulating the sum of $SOC_{i,t+1}^{top}$ and ΔSOC_{max}, we can derive the top $\hat{d}_{t+1}$ SOC of batteries available for EV replacement. The additional energy required to meet the predicted demand is calculated by subtracting the available energy in the BSS from the predicted desired energy amount. This energy need, when multiplied by the penalty weight β', results in the potential extra cost due to unfulfilled projected demand.

Consequently, the reward function accommodating STP can be updated as follows:

$$\mathcal{R}_{STP}(s_t, a_t) = -(C_{elec} + C_{demand} + C_{capacity} + C_{elec}^{STP} + C_{demand}^{STP}) \tag{3.20}$$

This function integrates the traditional costs of electricity C_{elec}, demand C_{demand}, and capacity $C_{capacity}$, with the additional costs introduced for STP, namely C_{elec}^{STP} and C_{demand}^{STP}.

Reward Function with LTP

To account for LTP, two additional components are introduced. The first component, C_{elec}^{LTP}, represents the potential additional electricity cost if we fail to take advantage of the lowest predicted electricity price over the next n time slots (p_{min}^{LTP}) for charging the EV batteries:

$$C_{elec}^{LTP} = \alpha'' \cdot \max(0, \sum_{m=1}^{M} \frac{a_{m,t}}{\eta} \cdot p_t \cdot \Delta t - \sum_{m=1}^{M} \frac{a_{m,t}}{\eta} \cdot p_{min}^{LTP} \cdot \Delta t) \tag{3.21}$$

$$p_{min}^{LTP} = \min_{t \leq k \leq t+n} p_k \tag{3.22}$$

In Eq. (3.21), C_{elec}^{LTP} is calculated by comparing the current electricity cost at price p_t with the cost at the predicted minimum electricity price p_{min}^{LTP} over the next n time slots. Eq. (3.22) identifies the minimum electricity price over the next n time slots. If the cost at p_{min}^{LTP} is lower than current cost, the model is encouraged to increase the charging speed, and vice versa. As a result, the RL agent is influenced to adjust the charging strategy in anticipation of future electricity price fluctuations.

$$C_{demand}^{LTP} = \beta'' \cdot \sum_{j=1}^{J} \sum_{i=0}^{\hat{d}_{t+j}} \max(0, [SOC_{t+j}^{\hat{Desire}} \cdot B - (SOC_{i,t+j}^{top} + \Delta SOC_{max}) \cdot B]) \quad (3.23)$$

$$SOC_{t+j}^{\hat{Desire}} = SOC_{ave}^{EV} + \frac{\hat{\boldsymbol{e}}_{t+j}^{ave}}{B} \cdot 100\% \quad (3.24)$$

$$SOC_{i,t+j}^{top} = SOC_{t+j}^{BSS,[i]} \quad \text{for} \quad i \in 0, 1, 2, ..., \hat{d}_{t+1}, \quad j \in 1, 2, ..., n \quad (3.25)$$

Eq. (3.23) introduces another component to account for LTP, which is C_{demand}^{LTP}. It represents the potential additional demand cost if the BSS fails to meet the expected demand of EV over the next J time slots. The cost is calculated by comparing the required energy amount for each expected demand with the available energy amount in the BSS. This component encourages the model to increase the charging speed when the BSS might not be able to meet the expected future demand. In more detail, β'' is a penalty weight that can be adjusted for LTP, $\hat{d}_{t+j}$ is the expected demand at time slot $t + j$, SOC_{ave}^{EV} is the average SOC of EVs, B is the battery capacity, $\hat{\boldsymbol{e}}_{t+j}^{ave}$ is the average energy requirement of EVs at time slot $t + j$, $SOC_{i,t+j}^{top}$ is the SOC of the i-th battery in the BSS at time slot $t + j$, and ΔSOC_{max} is the maximum possible increase in SOC during a time slot. Eq. (3.25) shows how $SOC_{i,t+j}^{top}$ is determined. It is simply the SOC of the i-th battery in the BSS at time slot $t + j$. This equation ensures that the model considers the SOC of the top $\hat{d}_{t+j}$ batteries in the BSS, which are the ones that are expected to be used to meet the demand. In this context, the model virtually "swaps" the batteries at the end of time slot based on the expected demand $\hat{d}_{t+j}$ and the average SOC of EV, SOC_{ave}^{EV}. The batteries are charged with the maximum charging speed, ΔSOC_{max}, to replenish the energy. This operation aims to simulate the potential situation and calculate the expected cost for the upcoming n time slots.

The final reward function incorporating LTP can now be written as:

$$\mathcal{R}_{LTP}(s_t, a_t) = -(C_{elec} + C_{demand} + C_{capacity} + C_{elec}^{LTP} + C_{demand}^{LTP}) \quad (3.26)$$

Normalized Reward Function

The normalization of the reward function components can aid in achieving more robust and stable training of the RL agent [51]. By ensuring that the individual components C_{elec} in Eq. (3.12), C_{demand} in Eq. (3.13), and $C_{capacity}$ in Eq. (3.15) fall within the range of 0 and 1, we not only facilitate the RL agent's understanding of the relative importance of the three components, but also prevent extreme values from adversely affecting the learning process.

$$C_{elec}^{'} = \alpha \cdot \frac{max(0, (\sum_{m=1}^{M} \frac{a_{m,t}}{\eta} \cdot p_t \cdot \Delta t - \sum_{m=1}^{M} \frac{a_{m,t}}{\eta} \cdot p_{min} \cdot \Delta t))}{\sum_{m=1}^{M} \frac{a_{m,t}}{\eta} \cdot (p_{max} - p_{min}) \cdot \Delta t} \tag{3.27}$$

$$C_{demand}^{'} = \beta \cdot \frac{\sum_{j=1}^{d_t} \max(0, SOC_{j,t}^{Desire} - SOC_{j,t+1}^{BSS}) \cdot B}{\sum_{j=1}^{d_t} SOC_{j,t}^{Desire} \cdot B} \tag{3.28}$$

$$C_{capacity}^{'} = \gamma \cdot \frac{\sum_{m=1}^{M} \max(0, SOC_{max} - SOC_{m,t+1}^{BSS}) \cdot B}{M \cdot B} \tag{3.29}$$

The normalization of $C_{elec}^{'}$ in Eq. (3.27) ensures that the RL agent considers the fluctuations in electricity price, optimizing the charging speed accordingly.

Similarly, the normalization of $C_{demand}^{'}$ in Eq. (3.28) ensures that the RL agent places proper emphasis on meeting the energy demands of EVs, with the normalized value indicating the proportion of unmet demand to the total demand.

The normalization of $C_{capacity}^{'}$ in Eq. (3.29) indicates the proportion of unused battery capacity to the total capacity, encouraging the RL agent to maintain the state of charge close to its maximum capacity.

Then the normalized reward function can be written as follows:

$$\mathcal{R}^{'}(s_t, a_t) = -(C_{elec}^{'} + C_{demand}^{'} + C_{capacity}^{'}) \tag{3.30}$$

represents a balance between minimizing electricity costs, meeting EV demands, and maintaining battery capacity.

Two additional normalized reward functions are introduced after incorporating STP. To incorporate the range of values within 0 to 1 for Eq. (3.16), we can restructure the equation as follows:

$$C_{elec}^{'STP} = \alpha^{'} \cdot \frac{\max(0, \sum_{m=1}^{M} \frac{a_{m,t}}{\eta} \cdot p_t \cdot \Delta t - \sum_{m=1}^{M} \frac{a_{m,t}}{\eta} \cdot p_{t+1} \cdot \Delta t)}{\sum_{m=1}^{M} \frac{a_{m,t}}{\eta} \cdot (p_{max} - p_{min}) \cdot \Delta t} \tag{3.31}$$

In this normalized version of the equation, Eq. (3.31), the denominator represents the maximum possible cost variation for the given charging speeds, which occurs when the electricity price shifts from p_{min} to p_{max}.

To normalize Eq. (3.17), we divide the total unmet demand by the total predicted demand, yielding:

$$C'^{STP}_{demand} = \beta' \cdot \frac{\sum_{i=1}^{\hat{d}_{t+1}} \max(0, \widehat{SOC^{Desire}_{t+1}} \cdot B - (SOC^{top}_{i,t+1} + \Delta SOC_{max}) \cdot B)}{\hat{d}_{t+1} \cdot (SOC^{EV}_{ave} \cdot B + \hat{\boldsymbol{e}}^{ave}_{t+1})} \tag{3.32}$$

In this normalized form of the equation, Eq. (3.32), the denominator is the predicted total desired electricity amount for the next time slot. This normalization ensures that C'^{STP}_{demand} falls between 0 and 1.

The reward function accounting for STP can be rewritten as follows:

$$\mathcal{R}'_{STP}(s_t, a_t) = -(C'_{elec} + C'_{demand} + C'_{capacity} + C'^{STP}_{elec} + C'^{STP}_{demand}) \tag{3.33}$$

balances minimizing electricity costs, meeting EV demands, and maintaining battery capacity, in addition to considering predicted electricity costs and demand.

To ensure the normalization of the LTP embedded reward function components within the range of 0 to 1, we reformulate Eq. (3.21) and Eq. (3.23) as follows:

$$C'^{LTP}_{elec} = \alpha'' \cdot \frac{\max(0, \sum_{m=1}^{M} \frac{a_{m,t}}{\eta} \cdot p_t \cdot \Delta t - \sum_{m=1}^{M} \frac{a_{m,t}}{\eta} \cdot p^{LTP}_{min} \cdot \Delta t)}{\sum_{m=1}^{M} \frac{a_{m,t}}{\eta} \cdot (p_{max} - p_{min}) \cdot \Delta t} \tag{3.34}$$

$$C'^{LTP}_{demand} = \beta'' \cdot \frac{\sum_{j=0}^{J} \sum_{i=1}^{\hat{d}_{t+j}} \max(0, [\widehat{SOC^{Desire}_{t+j}} \cdot B - (SOC^{top}_{i,t+j} + \Delta SOC_{max}) \cdot B])}{\sum_{j=0}^{n} [\hat{d}_{t+j} \cdot (SOC^{EV}_{ave} \cdot B + \hat{\boldsymbol{e}}^{ave}_{t+j})]} \tag{3.35}$$

In Eq. (3.34), the denominator represents the maximum possible cost variation for the given charging speeds, which occurs when the electricity price shifts from p_{min} to p_{max}.

In Eq. (3.35), the denominator is the sum of predicted total desired electricity amount for the next J time slots.

Finally, we can express the reward function with LTP components as follows:

$$\mathcal{R}(s_t, a_t) = -(C'_{elec} + C'_{demand} + C'_{capacity} + C'^{LTP}_{elec} + C'^{LTP}_{demand}) \tag{3.36}$$

This reward function balances minimizing electricity costs, meeting EV demands, and maintaining battery capacity, while also considering long-term predictions of electricity costs and demand.

Algorithm 4 Deep Deterministic Policy Gradient

```
Initialize actor network $\theta_\mu$ and its target network $\theta_{\mu'}$
Initialize critic network $\theta_Q$ and its target network $\theta_{Q'}$
Initialize replay buffer $M$
while not converged do
    for each episode $t$ = 1, 2, 3, ... do
        Observe state $s_t$
        Select action $a_t = \mu_{\theta_\mu}(s_t) + \epsilon$, with $\epsilon \sim \mathcal{N}(0, \sigma)$
        Execute action $a_t$ and observe reward $r_t$ and new state $s_{t+1}$
        Store transition tuple $(s_t, a_t, r_t, s_{t+1})$ in $M$
        Sample a mini-batch of transitions from $M$
        Update critic by minimizing the loss:
        $L = \frac{1}{N}\sum_i (y_i - Q_{\theta_Q}(s_i, a_i))^2$
        with $y_i = r_i + \gamma Q_{\theta_{Q'}}(s_{i+1}, \mu_{\theta_{\mu'}}(s_{i+1}))$
        Update the actor policy using the sampled policy gradient:
        $\nabla_{\theta_\mu} J \approx \frac{1}{N}\sum_i \nabla_a Q_{\theta_Q}(s, a)|s = s_i, a = \mu\theta_\mu(s_i)\nabla_{\theta_\mu}\mu_{\theta_\mu}(s)|s_i$
        Update target networks:
        $\theta\mu' = \tau\theta_\mu + (1-\tau)\theta_{\mu'}$
        $\theta_{Q'} = \tau\theta_Q + (1-\tau)\theta_{Q'}$
    end for
end while
```

3.2.3 Utilizing Reinforcement Learning for Charging Optimization

Our study leverages the DDPG algorithm [37], a renowned model-free, off-policy actor-critic algorithm adept at addressing continuous control challenges. This algorithm is particularly effective in optimizing high-dimensional action or policy spaces and employs four neural networks: the actor network, actor target network, critic network, and critic target network. The actor network is responsible for formulating policy decisions, acting as a decoder that interprets system states and develops appropriate policies. On the other hand, the critic network functions as an evaluator of the policies proposed by the actor network, taking in a blend of observations and policies and utilizing the reward value as a performance indicator. Equipped with the knowledge gained from the critic network, the actor network can then formulate policies that lead to higher rewards. The actor and critic target networks are incorporated to enhance learning stability by gradually adjusting the values within the network. This results in the DDPG algorithm learning to adapt charging rates in accordance with resource availability. The learning process involves continuous optimization of the decision-making approach, modifying the charging rates based on the prevailing system state, resource availability, and lessons learned from past experiences. The actor network's focus on optimal policy creation, coupled with the critic network's emphasis on high reward generation, fosters a culture of continuous progress within the algorithm. Operating within a reinforcement learning framework, the actor-critic duo learns the most effective strategies through a process of trial and error, using feedback from the environment to hone its decisions. The inclusion of target networks further steadies this learning process, providing a more consistent target for learning updates and reducing the risk of destabilizing feedback loops, a common issue in traditional reinforcement learning methods. This stable learning environment facilitates more precise and reliable policy optimization, ensuring that the DDPG algorithm consistently delivers decisions that boost the performance of the BSS.

In this chapter, we introduce two proposed schemes, RLC-S and RLC-L. Both of these schemes incorporate a prediction module to anticipate future trends and demands, playing a crucial role in improving the effectiveness of the BSS. Specifically, RLC-S integrates a STP, focusing on immediate future demands, while RLC-L integrates a LTP, focusing on broader and longer-term trends. We expect both of these schemes to outperform existing strategies, offering enhanced predictive accuracy and operational efficiency in the realm of BSS. Further details and performance evaluations of these schemes will be discussed in the following chapters.

Chapter 4

Performance Evaluation

In this chapter, we provide a comprehensive overview of the experiments conducted to evaluate the performance of our proposed DRL-based BSS optimization framework. We start with an explanation of the dataset used in the study, followed by the experimental setup and the evaluation metrics we employed to assess the performance of our approach. Lastly, we present the results obtained from these experiments and compare them to other relevant benchmarks and methods.

Current charging/swapping schemes cannot be directly utilized, which is why we did not include state-of-the-art schemes in our experiments. Instead, we focused our experiments on comparing RLC-S and RLC-L with the following charging methods:

- Greedy Charging: A policy that sets a fixed maximum charging speed irrespective of the demand or electricity price.

- Random Charging: This policy randomly selects battery charging rates within a defined minimum and maximum range.

- Demand-based Charging (DBC): This algorithm charges N batteries at full speed when N vehicles arrive at a given timeslot, while other batteries are charged at half the maximum speed.

- DDPG-based Charging: This policy uses DDPG to determine the charging rates. However, no predictive information is utilized by DDPG.

4.1 Experimental Dataset

Selecting an appropriate dataset that accurately represents the arrival patterns of EVs and the operational conditions of a real-world BSS is important to effectively evaluate the performance of our proposed RLC scheme. Previous studies have mainly adopted two approaches for obtaining such data: simulated data and EV charging databases.

Simulated data, generated through simulations, allows for controlled experiments and the manipulation of specific variables. However, it has limitations in fully capturing the complexity and variability of real-world scenarios, which may affect the generalizability of the experimental results. On the other hand, data extracted from existing EV charging databases offers the advantage of working with real-world information. However, this data requires adaptation and preprocessing to fit the context of a BSS, as the operational requirements and customer interactions in battery swapping systems may differ from conventional charging stations.

Table 4.1: Electricity Rates

TOU Price Periods	**TOU Prices (¢/kWh)**
Off-Peak 7 p.m. – 7 a.m.	7.4
Mid-Peak 11 a.m. – 5 p.m.	10.2
On-Peak 7 a.m. – 11 a.m. and 5 p.m. – 7 p.m.	15.1

In our research, we opted to use data derived from an EV charging database, ACN data [8], to ensure that our experiments are grounded in real-world conditions. EV battery demand numbers and EV battery swapping load are extracted from the ACN dataset. This dataset provides information on EV arrivals and corresponding energy requests. Additionally, we integrated the Time of Use (TOU) electricity rates defined by the Ontario Energy Board [52], as shown in Table 4.1. These choices allow us to better evaluate the effectiveness of the RLC-S and RLC-L schemes in comparison to alternative charging schemes and to demonstrate their ability to address the unique challenges associated with BSS management. Moreover, working with real-world data contributes to the relevance and generalizability of our findings, thereby increasing the practical applicability of our proposed RLC scheme.

4.2 Experiment Configuration

The scale of a BSS depends on the number of batteries it contains. In our simulations, we considered three different settings with varying scales. Specifically, we examined a BSS equipped with 15 batteries, another with 10 batteries, and a third with 5 batteries. The specific numbers allow for the examination of small (5 batteries), medium (10 batteries), and relatively large (15 batteries) scale operations. This range can help understand how different scales impact the performance of the proposed schemes. In addition, these numbers are reasonably realistic for real-world BSS, making the outcomes of the simulation more applicable and useful [11]. Unless otherwise specified, the remaining parameters were kept consistent across all settings.

The MDP time slots were defined with time discretized into quarters [53][24][23]. The charging efficiency η was set to 0.9 [24], and each battery B had a capacity of 50 kWh. The maximum and minimum charging rates, $P_{\max}$ and $P_{\min}$, were set to 50 kW and 0 kW, respectively [54]. For the arriving EVs' SOC, $SOC_{j,t}^{EV}$, we used a Gaussian distribution with a mean of 30% and a standard deviation of 5% [55]. As mentioned in the previous section, we extracted the number of demanded batteries d_t, the electricity load for each EV e_t, and the TOU electricity price p_t from real-world data.

We used 65 weeks of data to train both the ensemble learning-based predictive model and the DDPG agent, while 5 weeks of data were employed to evaluate the performance of the proposed RLC-S and RLC-L schemes. The DDPG agent utilized a rectified linear unit (ReLU) activation function for the critic networks and a hyperbolic tangent (tanh) activation function for the actor networks. The discount factor was set to 0.99. The replay buffer size was set to 100,000, the minimum batch size was set to 1,000, and the batch size for learning was 64. All of the base models of ensemble learning were configured with a hidden layer size of 128. The learning rate was set to 0.01, and we trained the models using the Adam optimizer. The loss function was defined as the mean squared error between predicted and actual values. A look-back window of 96 time slots, equivalent to one day, and 192 time slots, equivalent to two days, was used for STP and LTP, respectively, based on the following experimental results. We assume that the threshold θ is 10%. This threshold is considered as a compromise between the user's energy demand and the BSS's available resources. The value of 10% is chosen under the assumption that even if the BSS cannot fully meet an EV's energy demand, an EV user would likely accept a replacement battery that can provide at least 10% SOC, which should enable the EV to reach another charging station or continue its journey for a short distance.

4.3 Evaluation Metrics

We employed the following metrics to evaluate the performance of our proposed RLC-S and RLC-L schemes:

- **Electricity Cost:** This corresponds to the cumulative expense of electricity utilized by the BSS. A lower electricity cost indicates that the BSS is efficiently managing its energy consumption and making better use of dynamic pricing signals. Electricity costs are in logarithmic scale.

- **Average SOC Discrepancy Rate:** This metric indicates the mean proportional difference between the targeted and actual SoC levels received during a battery exchange. A lower average SOC discrepancy rate implies that the BSS is providing batteries with SoC levels closer to the desired levels, resulting in higher customer satisfaction and better operational performance.

- **Battery Service Rate:** This metric defines the proportion of the total number of batteries serviced by the BSS to the total count of EV battery demands. A higher battery service rate suggests that the BSS is effectively meeting the demand for battery swaps and efficiently utilizing BSS resources.

To assess the accuracy and effectiveness of the prediction module, we utilize two evaluation metrics:

- **Mean Absolute Error (MAE):** This metric measures the average absolute difference between the predicted and actual values for variables such as battery demand and electricity price.

- **Root Mean Squared Error (RMSE):** This metric calculates the square root of the average squared difference between the predicted and actual values for a variable.

4.4 Prediction Module Evaluation

We assessed the performance of our ensemble-based prediction model by comparing it to three base models: RNN, LSTM, and GRU, using the MAE and RMSE metrics.

4.4.1 Total Electricity Load Prediction

Fig. 4.1 presents a 24-hour sample data, showcasing the actual curve of the total electricity load along with the predicted results generated by different algorithms. As depicted in this graph, it is clear that all predicted algorithms successfully follow the trend of the true total electricity load curve. We observe that before 5:00, the total electricity load remains nearly 0 kWh. During the peak hour, from 6:00 to 8:00, the electricity load reaches a maximum of approximately 300 kWh. Besides this major peak, a smaller peak can be seen at 12:00 when the electricity load reaches around 100 kWh. Additionally, we identify a mild period of increased electricity load between 12:00 and 16:00, with a load of about 50 kWh.

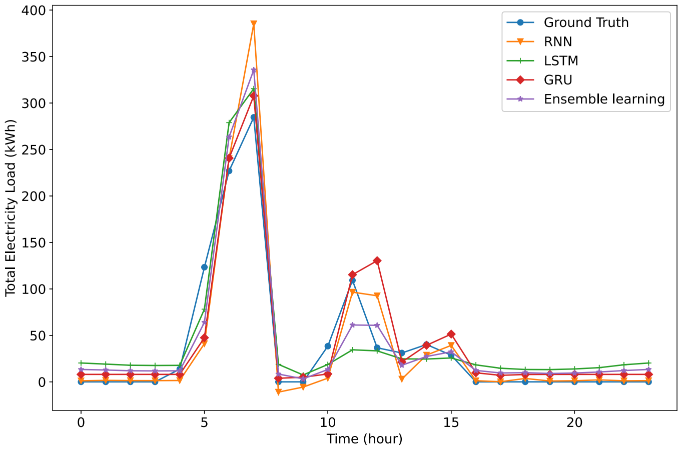

Figure 4.1: 24-hour Electricity Load Curve with Prediction Results

In our analysis, we further investigate a one-week electricity load profile, focusing solely on weekdays, as illustrated in Fig. 4.2. This graph displays the daily variations in electricity demand from Monday to Friday, showing that each prediction algorithm effectively follows the actual trend of the total electricity load. Concurrently, we observe recurring daily peaks and troughs throughout the weekdays, which are consistent with the patterns identified in the 24-hour total electricity load curve in Figure 4.1. Notably, the peak hour occurs around morning, and mild peak hours can be observed in the afternoon.

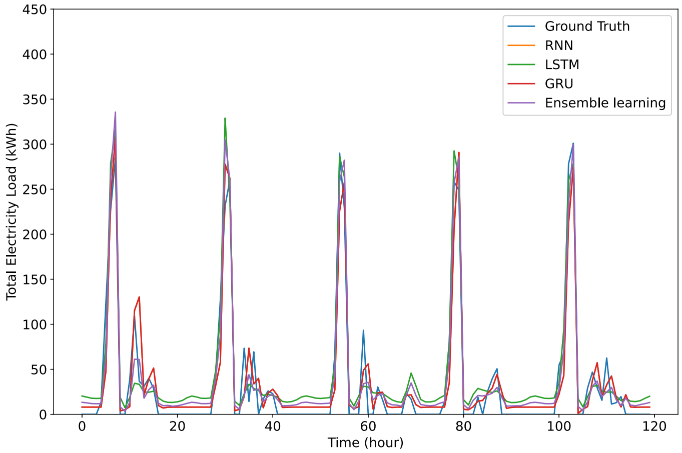

Figure 4.2: One Week Electricity Load Curve with Prediction Results

Total Electricity Load with STP

Table 4.2 presents a comparison of the Ensemble Learning model's performance with that of the three base models (RNN, LSTM, and GRU) using the MAE and RMSE metrics. The look-back time step is set to 96, which is equivalent to 24 hours, and the output is focused on STP with 1 predicted time step. As shown in the table, the Ensemble Learning model exhibits superior performance, achieving the lowest MAE of 7.232788 and RMSE of 17.460322. In terms of MAE, the Ensemble Learning model demonstrates improvements of approximately 14.16%, 7.25%, and 14.73% over the RNN, LSTM, and GRU models, respectively. Similarly, for the RMSE metric, the Ensemble Learning model displays enhancements of around 5.0%, 1.56%, and 5.95% compared to the RNN, LSTM, and GRU models, respectively. These results underline the effectiveness of the ensemble learning approach in predicting short-term electricity load with a 96 look-back time step and 1 predicted time step. The substantial percentage improvements achieved by the ensemble learning model further emphasize its superiority over the base models in terms of prediction accuracy.

To further examine the effectiveness of our ensemble learning-based model for STP of total electricity load, we set the look-back time step to 192, which is equivalent to 48 hours. The results can be seen in Table 4.3. With a 192 look-back time step, the Ensemble Learning model continues to outperform the base models, yielding the lowest MAE and RMSE values. In terms of MAE, the Ensemble Learning model shows improvements of approximately 11.12%, 3.83%, and 4.34% over the RNN, LSTM, and GRU models, respectively. Similarly, for the RMSE metric, the Ensemble Learning model demonstrates enhancements of around 5.66%, 0.73%, and 1.16% compared to the RNN, LSTM, and GRU models, respectively. These results highlight the robustness and adaptability of the ensemble learning approach in handling different look-back time steps, further establishing its effectiveness for STP of total electricity load.

Table 4.2: Total Electricity Load Prediction with 96 Look-back Time Step and 1 Predicted Time Step

Metric	**RNN**	**LSTM**	**GRU**	**Ensemble Learning**
MAE	8.425430	7.798631	8.482590	**7.232788**
RMSE	18.379395	17.737808	18.564583	**17.460322**

Table 4.3: Total Electricity Load Prediction with 192 Look-back Time Step and 1 Predicted Time Step

Metric	**RNN**	**LSTM**	**GRU**	**Ensemble Learning**
MAE	8.719349	8.148189	8.187557	**7.847148**
RMSE	18.503582	17.639829	17.714649	**17.511306**

Total Electricity Load with LTP

For LTP, Table 4.4 provides a performance assessment of the Ensemble Learning model compared to the three base models (RNN, LSTM, and GRU) using the MAE and RMSE metrics. The look-back time step is set to 96, equivalent to 24 hours, and the prediction focuses on 24 time steps. The Ensemble Learning model consistently outperforms the base models, achieving the lowest MAE and RMSE values. Specifically, the Ensemble Learning model demonstrates significant improvements over the

RNN, LSTM, and GRU models in both metrics, with 6.44%, 1.96%, and 4.82% improvement in MAE, and 4.78%, 0.68%, and 2.57% improvement in RMSE, respectively. These results highlight the effectiveness of the ensemble learning approach for long-term electricity load forecasting with a 96 look-back time step and 24 predicted time steps.

Table 4.5 presents a performance comparison between the Ensemble Learning model and the base models using a 192 look-back time step and 24 predicted time steps. The Ensemble Learning model maintains its superior performance, achieving the lowest MAE and RMSE values. Specifically, the Ensemble Learning model demonstrates improvements of approximately 2.49%, 2.71%, and 5.69% in MAE, and 0.63%, 4.18%, and 3.74% in RMSE over the RNN, LSTM, and GRU models, respectively.

Table 4.4: Total Electricity Load Prediction with 96 Look-back Time Step and 24 Predicted Time Step

Metric	**RNN**	**LSTM**	**GRU**	**Ensemble Learning**
MAE	8.874624	8.468581	8.723219	**8.302767**
RMSE	20.063794	19.235966	19.608810	**19.104075**

Table 4.5: Total Electricity Load Prediction with 192 Look-back Time Step and 24 Predicted Time Step

Metric	**RNN**	**LSTM**	**GRU**	**Ensemble Learning**
MAE	8.394177	8.413154	8.680063	**8.185546**
RMSE	19.025359	19.729267	19.641021	**18.905046**

4.4.2 Number of Demanded Batteries

Fig. 4.3 and Fig. 4.4 depict the number of demanded batteries for a 24-hour period and a one-week period, respectively. As with the total electricity load prediction, the ensemble learning model, along with the base models (RNN, LSTM, and GRU), accurately follows the trend of the true number of demanded batteries. The demand pattern for batteries closely mirrors the electricity load pattern since the total electricity load is directly related to the energy required for battery demand.

In the 24-hour number of demanded batteries curve (Fig. 4.3), all prediction algorithms effectively capture the actual trend of the number of demanded batteries.

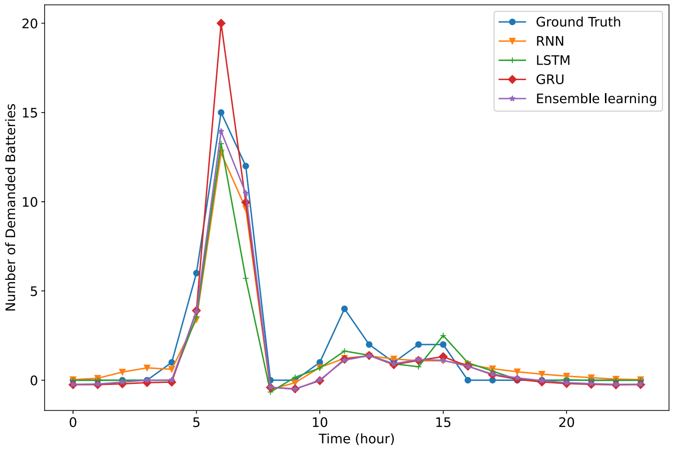

Figure 4.3: 24-hour Number of Demanded Batteries Curve with Prediction Results

We observe that battery demand is virtually non-existent before 5:00. The demand then experiences a significant increase during peak hours, from 6:00 to 8:00. A smaller peak is evident at 12:00, and a moderate rise in demand occurs between 14:00 and 16:00.

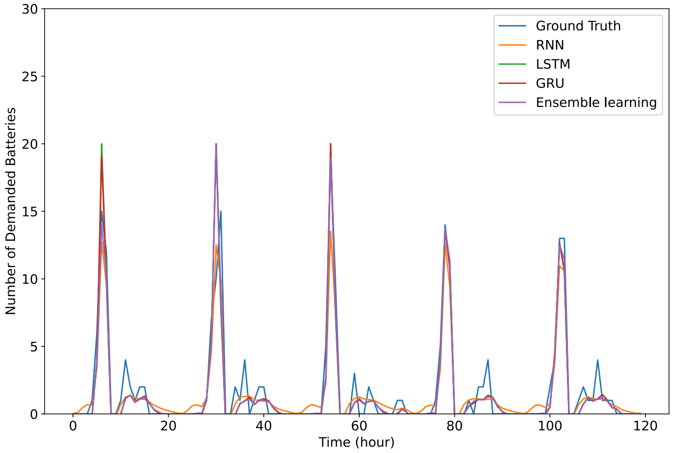

Figure 4.4: One week Number of Demanded Batteries Curve with Prediction Results

The one-week number of demanded batteries curve (Fig. 4.4) showcases patterns similar to the 24-hour curve, with daily fluctuations from Monday to Friday. The prediction algorithms effectively capture the actual trend of the number of demanded batteries. The peak hour consistently takes place around morning, with milder peak hours occurring in the afternoon.

Number of Demanded Batteries with STP

Table 4.6 presents a comparison of the prediction performance for the number of demanded batteries using the Ensemble Learning model and the three base models (RNN, LSTM, and GRU). The performance is evaluated based on two metrics: MAE and RMSE. The look-back time step is set to 96, which is equivalent to 24 hours, and the output focuses on STP with 1 time step of output. As shown in the table, the Ensemble Learning model outperforms the base models, achieving the lowest MAE of 0.265075 and RMSE of 0.713261. In terms of MAE, the Ensemble Learning model demonstrates improvements of approximately 25.36%, 17.45%, and 21.82% over the RNN, LSTM, and GRU models, respectively. Similarly, for the RMSE metric, the Ensemble Learning model displays enhancements of around 10.21%, 1.33%, and 1.53% compared to the RNN, LSTM, and GRU models, respectively. These results indicate that the Ensemble Learning model is more accurate in predicting the number of demanded batteries with a 96 look-back time step and a single predicted time step compared to the RNN, LSTM, and GRU models.

To further examine the effectiveness of our ensemble learning model, we set a 192 look-back time step and 1 predicted time step. As shown in Table 4.7, the Ensemble Learning model outperforms the base models, achieving the lowest MAE of 0.288359 and RMSE of 0.716058. The improvements in MAE are approximately 22.38%, 20.43%, and 30.09% over the RNN, LSTM, and GRU models, respectively. For the RMSE metric, the enhancements are around 5.87%, 1.00%, and 10.57% compared to the RNN, LSTM, and GRU models, respectively.

Table 4.6: Number of Demanded Batteries Prediction with 96 Look-back Time Step and 1 Predicted Time Step

Metric	**RNN**	**LSTM**	**GRU**	**Ensemble Learning**
MAE	0.355282	0.321016	0.339137	**0.265075**
RMSE	0.794378	0.722892	0.724356	**0.713261**

Table 4.7: Number of Demanded Batteries Prediction with 192 Look-back Time Step and 1 Predicted Time Step

Metric	**RNN**	**LSTM**	**GRU**	**Ensemble Learning**
MAE	0.371670	0.362359	0.412517	**0.288359**
RMSE	0.760759	0.723310	0.800816	**0.716058**

Number of Demanded Batteries with LTP

Table 4.8 provides a performance comparison of the Ensemble Learning model and the base models (RNN, LSTM, and GRU) for LTP of the number of demanded batteries. The look-back time step is set to 96, equivalent to 24 hours, and the output focuses on 24 predicted time step. As shown in the table, the Ensemble Learning model demonstrates superior performance, achieving the lowest MAE of 0.314241 and RMSE of 0.817748. The improvements in MAE are approximately 23.7%, 17.1%, and 21.0% over the RNN, LSTM, and GRU models, respectively. For the RMSE metric, the enhancements are around 5.73%, 0.36%, and 2.42% compared to the RNN, LSTM, and GRU models, respectively.

For a 192 look-back time step and 24 predicted time step, Table 4.9 presents the performance comparison of the Ensemble Learning model and the base models. The Ensemble Learning model continues to outperform the base models, achieving the lowest MAE of 0.307174 and RMSE of 0.809523. The improvements in MAE are approximately 25.1%, 15.4%, and 20.7% over the RNN, LSTM, and GRU models, respectively. For the RMSE metric, the enhancements are around 7.36%, 1.65%, and 2.42% compared to the RNN, LSTM, and GRU models, respectively.

Table 4.8: Number of Demanded Batteries Prediction with 96 Look-back Time Step and 24 Predicted Time Step

Metric	**RNN**	**LSTM**	**GRU**	**Ensemble Learning**
MAE	0.411739	0.379010	0.397670	**0.314241**
RMSE	0.867443	0.820693	0.837975	**0.817748**

4.4.3 Insights on the Prediction Module

In the previous sections, we compared the performance of Ensemble Learning with the base models (RNN, LSTM, and GRU) for both total electricity load prediction and

Table 4.9: Number of Demanded Batteries Prediction with 192 Look-back Time Step and 24 Predicted Time Step

Metric	**RNN**	**LSTM**	**GRU**	**Ensemble Learning**
MAE	0.410289	0.363161	0.387423	**0.307174**
RMSE	0.873763	0.823098	0.829572	**0.809523**

the number of demanded batteries prediction. The results reveal that the Ensemble Learning model consistently outperforms the base models in both scenarios.

For STP, when the look-back time step is set to 96, the performance of the Ensemble Learning model is better than the case when the look-back time step is set to 192. Conversely, for LTP, we observe that a 192 look-back time step yields better performance than a 96 look-back time step. Based on these findings, we will adopt a 96 look-back time step for STP and a 192 look-back time step for LTP in our forecasting model.

4.5 RLC Evaluation

We evaluate and compare the performance of our proposed RLC-S and RLC-L schemes with Greedy Charging, Random Charging and DDPG-based Charging schemes in the context of a BSS with 15 batteries, 10 batteries and 5 batteries respectively. We examine three essential metrics: electricity cost, average SOC discrepancy rate, and battery service rate.

4.5.1 Performance of RLC for BSS with 15 Batteries

Electricity Cost

As demonstrated in Figure 4.5, the logarithmic-scale representation of the electricity costs for the five charging strategies, the superior performance of both the RLC-S and RLC-L schemes is evident when compared to the Greedy Charging, Random Charging, DBC and DDPG - based Charging schemes. The RLC-L scheme notably outshines the others, registering the lowest electricity cost across all strategies. This translates to cost reductions of 89.05%, 67.26%, 75.81%, 13.4%, and 6.1% when compared to the Greedy Charging, Random Charging, DBC, DDPG - based Charging, and RLC-S schemes respectively.

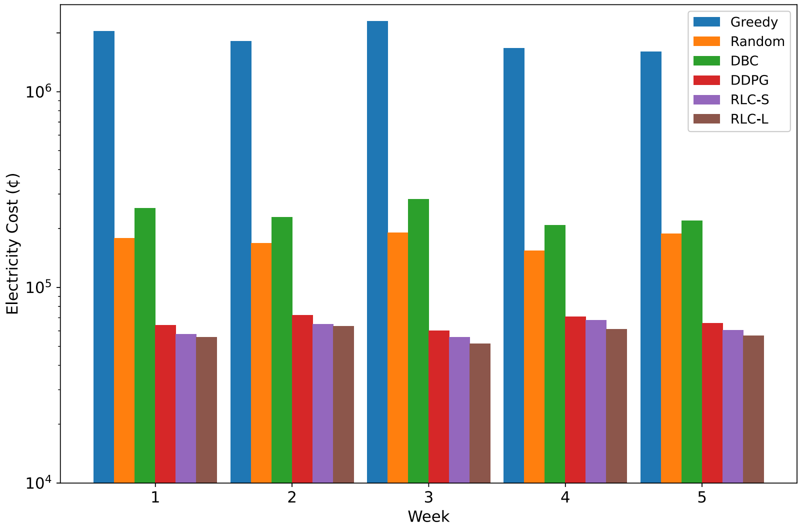

Figure 4.5: Electricity Cost for BSS with 15 Batteries

Average SOC Discrepancy Rate

The average SOC discrepancy rate for each strategy is shown in Fig. 4.6. A lower SOC discrepancy rate indicates that the BSS can supply batteries with SOC levels more in line with the customers' needs, improving customer satisfaction and the BSS's operational efficiency. In maintaining a lower average SOC discrepancy rate, the RLC-S and RLC-L schemes excel compared to the Random Charging, DBC and DDPG-based Charging schemes. The Greedy Charging strategy serves as a benchmark, due to its methodology of filling the battery in the BSS to the maximum possible extent. The RLC-L scheme proves to be superior by achieving the lowest average SOC discrepancy rate among the Random Charging, DBC, DDPG-based Charging, and RLC-S strategies, while closely matching the benchmark set by the Greedy Charging strategy. This result validates the RLC-L's ability to make accurate battery demand forecasts and adjust the charging strategies accordingly, thus catering to customer needs more effectively.

Battery Service Rate

Fig. 4.7 illustrates the battery service rates for the tested strategies. A higher battery service rate indicates a more efficient use of BSS resources and a greater ability to meet the demand for battery swaps. The RLC-S and RLC-L outperform the Random

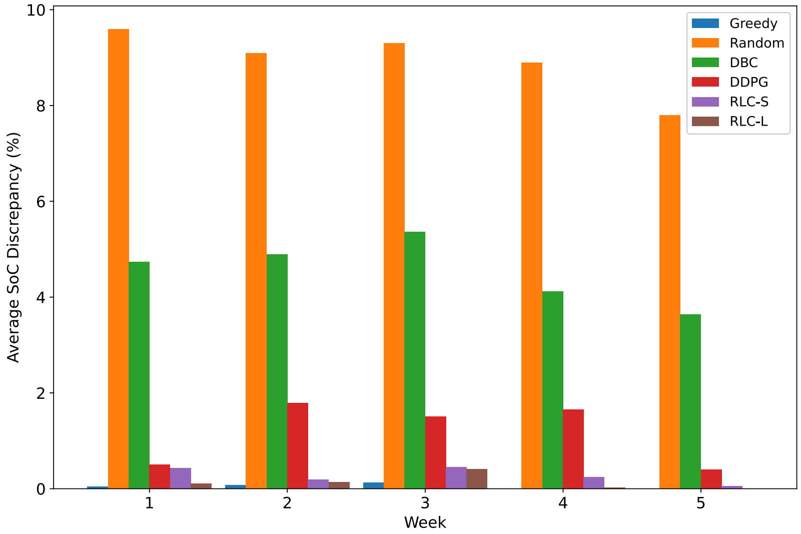

Figure 4.6: Average SOC Discreacy Rate for BSS with 15 Batteries

Charging, DBC and DDPG-based Charging schemes significantly in terms of battery service rate. The Greedy algorithm, again, serves as the benchmark for this metric. The RLC-L attains the highest battery service rate among all the strategies, coming close to the benchmark set by the Greedy Charging strategy. This highlights the RLC-L's proficiency in accurately predicting battery demand and effectively managing the BSS's battery inventory. Consequently, the RLC-L scheme is better equipped to fulfill battery demand and enhance the operational efficiency of the BSS.

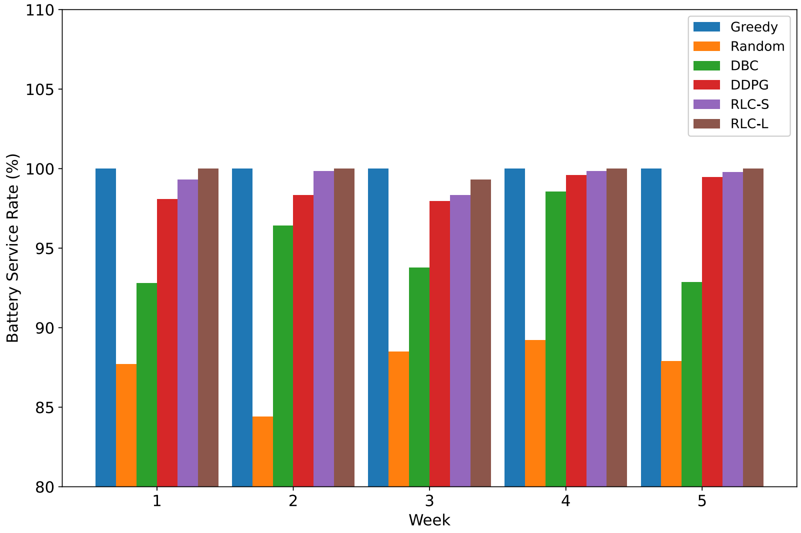

Figure 4.7: Battery Service Rate for BSS with 15 Batteries

4.5.2 Performance of RLC for BSS with 10 Batteries

Electricity Cost

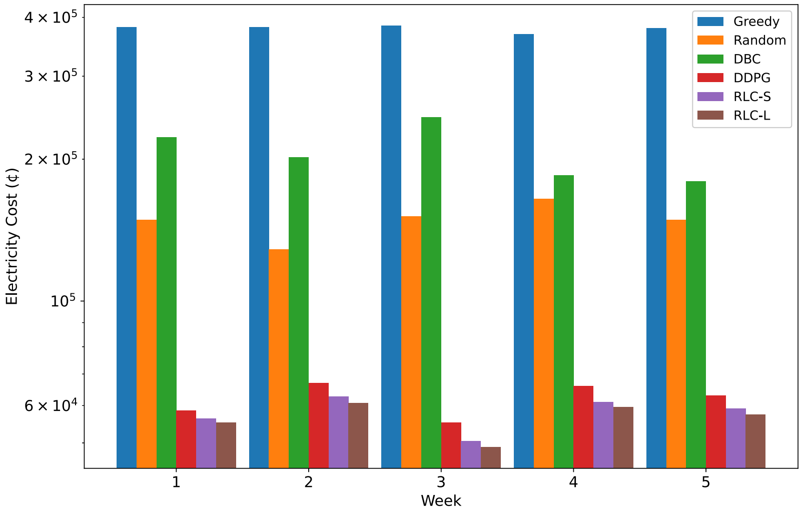

Figure 4.8: Electricity Cost for BSS with 10 Batteries

As illustrated in Fig. 4.8, both the RLC-S and RLC-L schemes continue to demonstrate superior performance over the baseline strategies in terms of electricity cost. Particularly, the RLC-L scheme achieves the lowest cost among all strategies. In more specific terms, the RLC-L strategy leads to cost reductions of 85.1%, 62.4%, 72.76%, 8.9%, and 2.7% when compared to the Greedy Charging, Random Charging, DBC, DDPG - based Charging, and RLC-S strategies, respectively.

Average SOC Discrepancy Rate

As depicted in Fig. 4.9, the RLC-S and RLC-L strategies maintain a superior performance in terms of achieving a lower average SOC discrepancy rate compared to the Random Charging, DBC and DDPG-based Charging methods, mirroring the results observed in the experiment with 15 batteries. The Greedy Charging strategy serves as a benchmark as well. Among these, the RLC-L strategy emerges as the most effective, achieving the lowest average SOC discrepancy rate, while closely matching the benchmark set by the Greedy Charging strategy. Notably, the performance of the RLC-L strategy is close to that of the upper bound set by the Greedy Charging strategy.

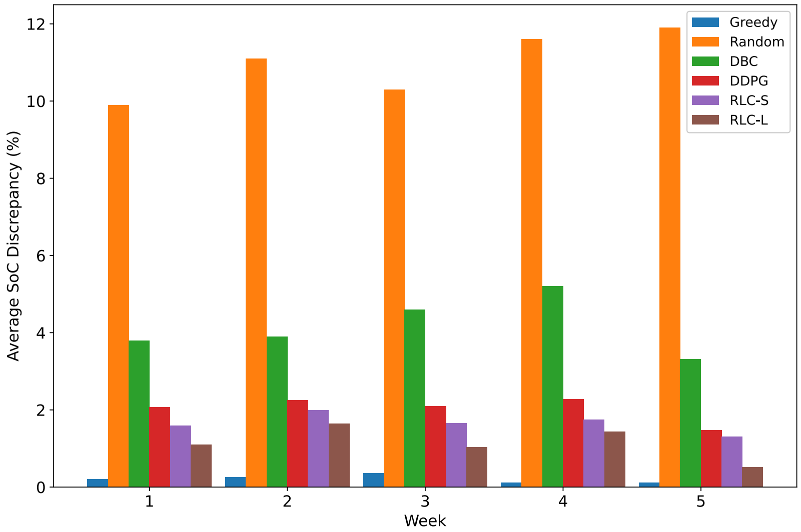

Figure 4.9: Average SOC Discrepancy Rate for BSS with 10 Batteries

Battery Service Rate

In the context of battery service rate, as depicted in Fig. 4.10, the RLC-S and RLC-L strategies significantly outperform Random Charging, DBC and DDPG-based Charging. The RLC-L scheme, in particular, excels by delivering the highest service rate, demonstrating its enhanced capacity to cater to battery demand and manage battery inventory effectively. Notably, the performance of the RLC - L scheme approaches the upper bound set by the Greedy Charging strategy.

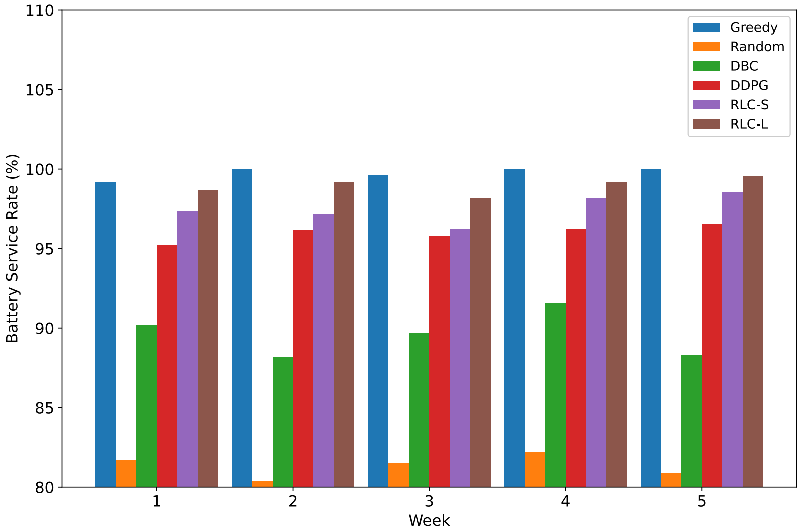

Figure 4.10: Battery Service Rate for BSS with 10 Batteries

4.5.3 Performance of RLC for BSS with 5 Batteries

Electricity Cost

As indicated in Fig. 4.11, both RLC-S and RLC-L schemes continue to showcase their superiority in managing electricity costs, even in the context of a smaller BSS. The RLC-L scheme shines once again, demonstrating the lowest electricity cost among all schemes. Specifically, the RLC-L strategy represents a cost reduction of 81.5%, 62.19%, 53.19%, 28.6% and 8.2% when compared to the Greedy, Random, DBC, DDPG, and RLC-S schemes, respectively. This consistent performance underlines the efficacy of our proposed schemes in optimizing electricity costs across different scales of BSS.

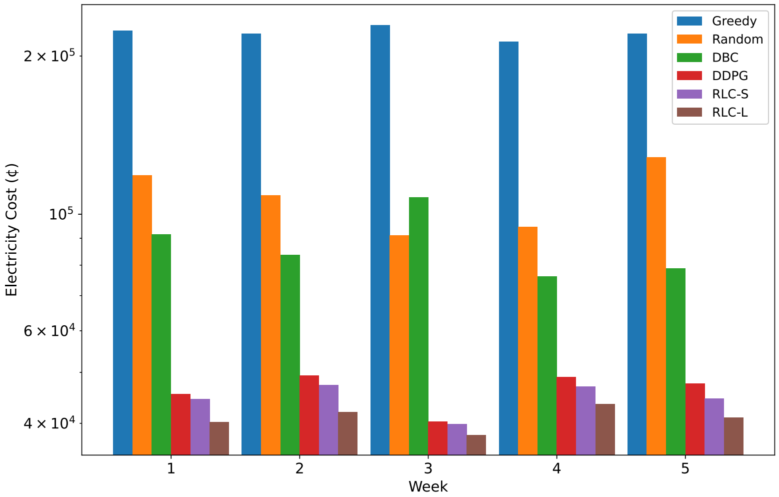

Figure 4.11: Electricity Cost for BSS with 5 Batteries

Average SOC Discrepancy Rate

As depicted in Fig. 4.12, the RLC-S and RLC-L schemes continue to outperform the Random Charging, DBC and DDPG-based Charging strategies in terms of maintaining a lower average SOC discrepancy rate. The RLC-L scheme proves to be exceptional, securing the lowest average SOC discrepancy rate among all the strategies, and closely approximating the upper bound set by the Greedy Charging scheme. This reaffirms the superior capability of the RLC-L scheme in satisfying customer battery demands even within the confines of a smaller-scale BSS.

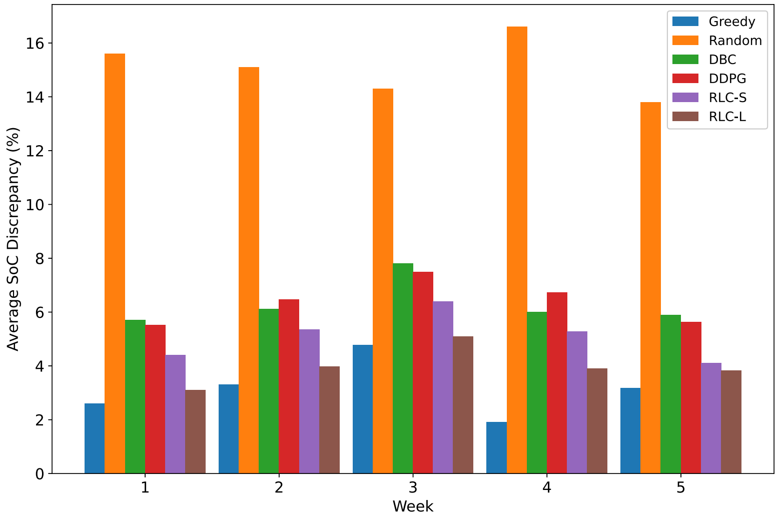

Figure 4.12: Average SOC Discrepancy Rate for BSS with 5 Batteries

Battery Service Rate

Fig. 4.13 showcases the performance of the RLC-S and RLC-L schemes in terms of battery service rate, with the RLC-L strategy outshining all others by achieving the highest rate. The RLC-L scheme, in particular, is remarkable in securing the topmost battery service rate, nearing the upper bound set by the Greedy scheme. This underscores its superior capability in accurately forecasting battery demand and effectively managing the BSS's battery inventory, irrespective of the BSS's size.

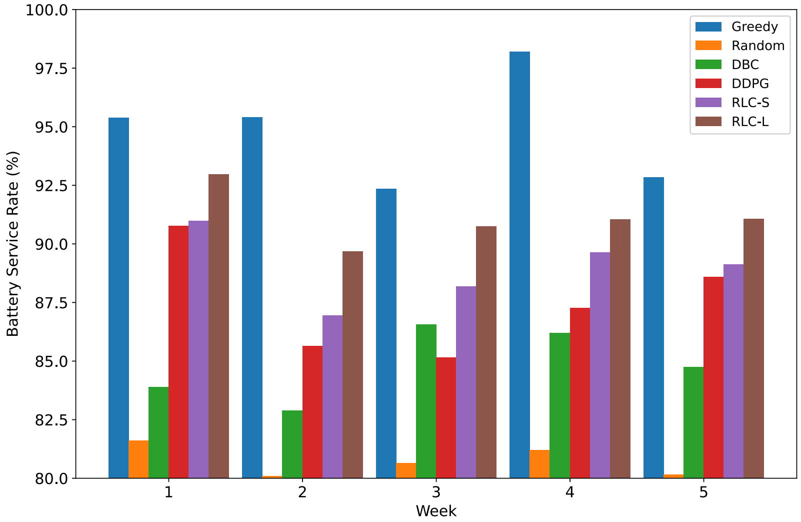

Figure 4.13: Battery Service Rate for BSS with 5 Batteries

4.5.4 Insights on RLC

Our results underscore the effectiveness of the proposed DRL-based BSS optimization framework, RLC, which incorporates predictive capabilities to guide the charging strategy. In particular, the RLC-L scheme, which uses long-term predictions, demonstrates superior performance across all metrics, highlighting the importance of foresight in BSS management.

Our RLC scheme can adeptly handle varying operational scales, demonstrating its robustness and scalability. Despite the decrease in available resources as the BSS size reduces from 15 to 5 batteries, the RLC maintains its superiority over the baseline methods in electricity cost, average SOC discrepancy rate, and battery service rate. This robustness suggests that the RLC scheme is capable of optimizing BSS operations across a range of different settings.

Additionally, the RLC scheme consistently outperforms the baseline methods, including the DDPG-based charging strategy that lacks predictive capabilities. This superiority underlines the value of incorporating predictive information into the DRL framework, with the RLC strategies utilizing projected trends in battery demand and electricity prices to make more informed decisions.

When it comes to customer satisfaction, as indicated by the average SOC discrepancy rate, the RLC scheme markedly outperforms the baseline methods. In particular, the RLC-L strategy secures the lowest average SOC discrepancy rate across all settings, implying that it is highly effective in meeting customers' preferences for battery SOC levels during a swap. This success is vital, given that meeting customer expectations is key to the commercial success of a BSS.

Furthermore, the RLC scheme excels in the battery service rate, demonstrating its proficiency in managing battery inventory to meet the demand for swaps. Notably, the RLC-L scheme consistently delivers the highest battery service rate, emphasizing its superior ability to balance the competing objectives of minimizing electricity cost and maximizing service level.

While our results highlight the efficacy of our RLC approach, they also showcase its significant improvement over existing solutions for BSS management. The key differences and advantages primarily lie in two aspects: scenario setup and methodological approach.

- Scenario: RLC addresses a dynamic real-world scenario where EVs arrive at BSS at different times with varying energy demands. By leveraging individual user profiles, including their energy demand and SOC, RLC personalizes the battery swapping process. This personalization ensures that each EV's energy demand is met as closely as possible, significantly enhancing the customer's battery swap experience.

- Approach: A defining characteristic of RLC is its integration of a prediction model into the DRL framework. By including short-term or long-term prediction into the charging strategy, RLC is capable of anticipating future EV arrivals and adjusting charging rates accordingly. This ability enables RLC to optimize the charging strategy in a way that minimizes costs while contributing to a more stable and efficient power grid.

Overall, RLC not only outperforms traditional methods in terms of operational efficiency but also in personalizing the battery swapping process based on individual EV requirements. By adopting a more realistic scenario setup and incorporating predictive capabilities into the DRL framework, our RLC approach is better equipped to address the complexities and challenges associated with BSS management.

Chapter 5

Conclusion and Future Work

5.1 Conclusion

Our research aimed to provide intelligent and practical solutions to the challenges involved in operating and managing BSS, a significant component of the rapidly expanding EV industry. The main objective of the proposed strategies was to markedly improve the operational efficiency of the BSS, while considering key variables such as the instability of electricity prices, the fluctuation of battery service rates, and varying SOC discrepancy rates.

To address these complexities, we introduced two novel methodologies, RLC-S and RLC-L. These schemes were designed with a keen emphasis on adaptability, scalability, and optimization, suitable for various scales of BSS operations. In practical situations, EVs dynamically arrive at BSS, each with unique energy needs and SOC. On arrival, the BSS operator receives each EV's profile. Accurately predicting future EV arrivals and profiles is crucial in deciding charging rates for each battery. If predictions show high energy-demand EVs arriving, the operator should proactively charge batteries at higher rates to meet demand. If only a few low-energy-demand EVs are expected, the operator might delay some charging. By effectively predicting future EV arrivals and adjusting charging rates accordingly, the operator can optimize the charging strategy, minimize costs, and contribute to a more stable and efficient power grid.

Our results indicate that both RLC-S and RLC-L consistently outperform traditional strategies like Greedy Charging, Random Charging, DBC and DDPG-based Charging. They demonstrated their effectiveness and efficiency by consistently achieving three primary objectives. First, they managed to lower electricity costs, which is vital given the instability of energy market prices. Second, they elevated battery service rates, indicating a more efficient usage of battery resources. Finally, they maintained a relatively low average SOC discrepancy rate, showing their capability to manage the difference between the charge levels of swapped out and in batteries.

However, while our findings are promising, it is vital to recognize certain limitations of our proposed methodologies. The models we presented are built upon key assumptions, such as all EVs being compatible with the BSS, and the battery swapping time being negligible in comparison to the charging time. Moreover, our methodologies currently do not account for potential battery degradation over time, which could considerably influence the BSS's efficiency and the batteries' lifespan. These assumptions represent potential areas for further investigation in future research.

Additionally, our prediction models, while generally accurate, are not flawless. Numerous real-world variables, including weather conditions, traffic patterns, and individual driving habits, can affect the arrival times and energy demands of EVs. These variables are not currently accounted for in our model, suggesting room for enhancement and fine-tuning to improve our BSS management strategies' effectiveness.

This thesis demonstrates how our methodologies not only excel beyond traditional methods, but also hold immense potential to transform the operational aspects of BSS in the EV industry. As we persist in our endeavor to create efficient, sustainable, and adaptable solutions, we are confident that these innovative strategies will play a significant role in the future of battery management within the EV landscape.

5.2 Future Work

The outcomes of this research present a wealth of opportunities for further investigations. Although the proposed RLC scheme has demonstrated significant effectiveness in managing BSS, several facets can be delved into for further refinement and optimization:

- **Incorporation of Additional Real-world Variables:** The simulations and experiments conducted in our study were based on a simplified representation of BSS operations. In practical scenarios, there are numerous other factors that could influence the performance of BSS. These include varying rates of battery degradation, a wide range of EV types, and diverse customer behaviors and preferences. By integrating these considerations into the RLC framework, we could potentially augment the model's precision and real-world relevance.

- **Implementation of Advanced Forecasting Models:** While our prediction module built on ensemble learning has yielded satisfactory results, the adoption of more sophisticated predictive models could enhance forecasting accuracy. For instance, transformer-based models or attention-based models, known for their efficiency in handling sequential data, might further improve the predictive power of our system.

- **Exploration of Multi-agent Learning:** Our research was based on the operation of a single BSS. In a real-world context, there are likely to be multiple BSS that interact and influence each other's operation. Delving into the realm of multi-agent reinforcement learning could furnish insights on how to optimally manage a network of interconnected BSS.

- **Integration with Power Grid Operations:** BSS also have the potential to be integrated into power grid operations, providing essential services such as frequency regulation and demand response. The development of a DRL framework that simultaneously optimizes both BSS operations and grid services could be a promising avenue of research, creating a more holistic solution for energy management in the context of EVs.

Overall, while our research has achieved significant progress in the management of BSS, these prospects for future work illustrate the vast potential for further exploration in this field, potentially leading to even more sophisticated and efficient strategies for BSS operation.

Bibliography

[1] International Energy Agency. *Global EV Outlook 2022: Securing supplies for an electric future.* OECD Publishing, Paris, 2022.

[2] Jidi Cao, Xin Chen, Rui Qiu, and Shuhua Hou. Electric vehicle industry sustainable development with a stakeholder engagement system. *Technology in Society*, 67:101771, 2021.

[3] G. Krishna. Understanding and identifying barriers to electric vehicle adoption through thematic analysis. *Transportation Research Interdisciplinary Perspectives*, 10:100364, 2021.

[4] Hao Wu. A survey of battery swapping stations for electric vehicles: Operation modes and decision scenarios. *IEEE Transactions on Intelligent Transportation Systems*, 23(8):10163–10185, 2022.

[5] Vedran Bobanac, Hrvoje Pandzic, and Tomislav Capuder. Survey on electric vehicles and battery swapping stations: Expectations of existing and future ev owners. In *2018 IEEE International Energy Conference (ENERGYCON)*, pages 1–6, 2018.

[6] Muhammad Shahid Mastoi, Shenxian Zhuang, Hafiz Mudassir Munir, Malik Haris, Mannan Hassan, Muhammad Usman, Syed Sabir Hussain Bukhari, and Jong-Suk Ro. An in-depth analysis of electric vehicle charging station infrastructure, policy implications, and future trends. *Energy Reports*, 8:11504–11529, 2022.

[7] Dingsong Cui, Zhenpo Wang, Peng Liu, Shuo Wang, David G. Dorrell, Xiaohui Li, and Weipeng Zhan. Operation optimization approaches of electric vehicle battery swapping and charging station: A literature review. *Energy*, 263:126095, 2023.

[8] Zachary J. Lee, Tongxin Li, and Steven H. Low. Acn-data: Analysis and applications of an open ev charging dataset. In *Proceedings of the Tenth ACM International Conference on Future Energy Systems*, e-Energy '19, page 139–149, New York, NY, USA, 2019. Association for Computing Machinery.

[9] Christopher Neuman, Andrew Meintz, and Myungsoo Jun. Workplace charging data collection and behavior. 11 2021.

[10] Rial A. Rajagukguk, Raden A. A. Ramadhan, and Hyun-Jin Lee. A review on deep learning models for forecasting time series data of solar irradiance and photovoltaic power. *Energies*, 13(24), 2020.

[11] Yusheng Zhang. Analysis of battery swapping technology for electric vehicles – using nio's battery swapping technology as an example. *SHS Web of Conferences*, 144:02015, 08 2022.

[12] Tianyang Zhang, Xi Chen, Zhe Yu, Xiaoyan Zhu, and Di Shi. A monte carlo simulation approach to evaluate service capacities of ev charging and battery swapping stations. *IEEE Transactions on Industrial Informatics*, 14(9):3914–3923, 2018.

[13] Saeed Salimi Amiri and Shahram Jadid. Optimal charging schedule of electric vehicles at battery swapping stations in a smart distribution network. In *2017 Smart Grid Conference (SGC)*, pages 1–8, 2017.

[14] Hao Wu, Grantham Kwok-Hung Pang, King Lun Choy, and Hoi Yan Lam. A charging-scheme decision model for electric vehicle battery swapping station using varied population evolutionary algorithms. *Applied Soft Computing*, 61:905–920, 2017.

[15] Luhao Wang and Massoud Pedram. Qos guaranteed online management of battery swapping station under dynamic energy pricing. *IET Cyber-Physical Systems: Theory & Applications*, 4(3):259–264, 2019.

[16] Qianwen Xu, Peng Wang, and Zhao Tiayang. Optimal operation of battery swapping-charging systems considering quality-of-service constraints. In *2017 IEEE Power and Energy Society General Meeting*, pages 1–5, 2017.

[17] Hao Wu, Grantham Kwok Hung Pang, King Lun Choy, and Hoi Yan Lam. An optimization model for electric vehicle battery charging at a battery swapping station. *IEEE Transactions on Vehicular Technology*, 67(2):881–895, 2018.

[18] Sarah G. Nurre, Russell Bent, Feng Pan, and Thomas C. Sharkey. Managing operations of plug-in hybrid electric vehicle (phev) exchange stations for use with a smart grid. *Energy Policy*, 67:364–377, 2014.

[19] Lijing Zhang, Suhua Lou, Yaowu Wu, Lin Yi, and Bin Hu. Optimal scheduling of electric vehicle battery swap station based on time-of-use pricing. In *2014 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, pages 1–6, 2014.

[20] Mohsen Mahoor, Zohreh S. Hosseini, and Amin Khodaei. Least-cost operation of a battery swapping station with random customer requests. *Energy*, 172:913–921, 2019.

[21] Weipeng Zhan, Zhenpo Wang, Lei Zhang, Peng Liu, Dingsong Cui, and David G. Dorrell. A review of siting, sizing, optimal scheduling, and cost-benefit analysis for battery swapping stations. *Energy*, 258:124723, 2022.

[22] Vidhya Murali, Abhik Banerjee, and Vijendran Gopalan Venkoparao. Optimal battery swapping operations using reinforcement learning. In *2019 Fifteenth International Conference on Information Processing (ICINPRO)*, pages 1–6, 2019.

[23] Yuan Gao, Jiajun Yang, Ming Yang, and Zhengshuo Li. Deep reinforcement learning based optimal schedule for a battery swapping station considering uncertainties. *IEEE Transactions on Industry Applications*, 56(5):5775–5784, 2020.

[24] Hang Luan, Xuefei Zhang, Jian Zhang, Qimei Cui, and Shuo Wang. A charging strategy with battery swapping station in car-sharing system using deep q-network. In *2021 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6, 2021.

[25] Min Seok Lee and Young Jae Jang. The agv battery swapping policy based on reinforcement learning. In *2022 IEEE 18th International Conference on Automation Science and Engineering (CASE)*, pages 1479–1484, 2022.

[26] Wang Qiang and Zhan Zhongli. Reinforcement learning model, algorithms and its application. In *2011 International Conference on Mechatronic Science, Electric Engineering and Computer (MEC)*, pages 1143–1146, 2011.

[27] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. A Bradford Book, The MIT Press, Cambridge, Massachusetts, London, England, 2nd edition, 2018.

[28] Beakcheol Jang, Myeonghwi Kim, Gaspard Harerimana, and Jong Wook Kim. Q-learning algorithms: A comprehensive classification and applications. *IEEE Access*, 7:133653–133667, 2019.

[29] Frank Emmert-Streib, Zhen Yang, Han Feng, Shailesh Tripathi, and Matthias Dehmer. An introductory review of deep learning for prediction models with big data. *Frontiers in Artificial Intelligence*, 3, 2020.

[30] Robin M. Schmidt. Recurrent neural networks (rnns): A gentle introduction and overview. *CoRR*, abs/1912.05911, 2019.

[31] Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. On the difficulty of training recurrent neural networks, 2013.

[32] Ralf C. Staudemeyer and Eric Rothstein Morris. Understanding LSTM - a tutorial into long short-term memory recurrent neural networks. *CoRR*, abs/1909.09586, 2019.

[33] Junyoung Chung, Çaglar Gülçehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *CoRR*, abs/1412.3555, 2014.

[34] Xu Wang, Sen Wang, Xingxing Liang, Dawei Zhao, Jincai Huang, Xin Xu, Bin Dai, and Qiguang Miao. Deep reinforcement learning: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–15, 2022.

[35] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. Playing atari with deep reinforcement learning, 2013.

[36] Richard S Sutton, David McAllester, Satinder Singh, and Yishay Mansour. Policy gradient methods for reinforcement learning with function approximation. In S. Solla, T. Leen, and K. Müller, editors, *Advances in Neural Information Processing Systems*, volume 12. MIT Press, 1999.

[37] Timothy P Lillicrap, Jonathan J Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*, 2015.

[38] Teng Liu, Xiaosong Hu, Shengbo Eben Li, and Dongpu Cao. Reinforcement learning optimized look-ahead energy management of a parallel hybrid electric vehicle. *IEEE/ASME Transactions on Mechatronics*, 22(4):1497–1507, 2017.

[39] Pengqian Yu, Joon Sern Lee, Ilya Kulyatin, Zekun Shi, and Sakyasingha Dasgupta. Model-based deep reinforcement learning for dynamic portfolio optimization, 2019.

[40] Clayton Miller, Pandarasamy Arjunan, Anjukan Kathirgamanathan, Chun Fu, Jonathan Roth, June Young Park, Chris Balbach, Krishnan Gowri, Zoltán Nagy, Anthony Fontanini, and Jeff Haberl. The ASHRAE great energy predictor III competition: Overview and results. *CoRR*, abs/2007.06933, 2020.

[41] Vladimir Popov, Mykola Fedosenko, Vadim Tkachenko, and Dmytro Yatsenko. Forecasting consumption of electrical energy using time series comprised of uncertain data. In *2019 IEEE 6th International Conference on Energy Smart Systems (ESS)*, pages 201–204, 2019.

[42] Da Huang. An optimized method for battery swapping demand prediction based on random forest regression. In *2021 IEEE 5th Information Technology,Networking,Electronic and Automation Control Conference (ITNEC)*, volume 5, pages 1739–1743, 2021.

[43] Alper Tokgöz and Gözde Ünal. A rnn based time series approach for forecasting turkish electricity load. In *2018 26th Signal Processing and Communications Applications Conference (SIU)*, pages 1–4, 2018.

[44] Shangfu Wei and Xiaoqing Bai. An attention-based cnn-gru model for resident load short-term forecast. In *2021 IEEE 5th Conference on Energy Internet and Energy System Integration (EI2)*, pages 2986–2991, 2021.

[45] Nakyoung Kim, Minkyung Kim, and Jun Kyun Choi. Lstm based short-term electricity consumption forecast with daily load profile sequences. In *2018 IEEE 7th Global Conference on Consumer Electronics (GCCE)*, pages 136–137, 2018.

[46] A. Okay Akyuz, Mitat Uysal, Berna Atak Bulbul, and M. Ozan Uysal. Ensemble approach for time series analysis in demand forecasting: Ensemble learning. In *2017 IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, pages 7–12, 2017.

[47] Thomas G. Dietterichl. Ensemble learning. In M. Arbib, editor, *The Handbook of Brain Theory and Neural Networks*, pages 405–408. MIT Press, 2002.

[48] Ludmila I Kuncheva and Christopher J Whitaker. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Machine learning*, 51(2):181, 2003.

[49] MICHAEL P. PERRONE and LEON N. COOPER. *When networks disagree: Ensemble methods for hybrid neural networks*, pages 342–358.

[50] Khushbu Kumari and Suniti Yadav. Linear regression analysis study. *Journal of the Practice of Cardiovascular Sciences*, 4:33, 01 2018.

[51] Tom Schaul, Georg Ostrovski, Iurii Kemaev, and Diana Borsa. Return-based scaling: Yet another normalisation trick for deep rl, 2021.

[52] Ontario Energy Board. Consumer information and protection, 2023.

[53] Mao Tan, Zhuocen Dai, Yongxin Su, Caixue Chen, Ling Wang, and Jie Chen. Bi-level optimization of charging scheduling of a battery swap station based on deep reinforcement learning. *Engineering Applications of Artificial Intelligence*, 118:105557, 2023.

[54] US Department of Transportation. Rural ev toolkit (pdf version). `https://www.transportation.gov/rural/ev/toolkit/pdf`, 2023. Accessed: 2023-05-10.

[55] Gurappa Battapothula, Chandrasekhar Yammani, and Sydulu Maheswarapu. Multi-objective optimal scheduling of electric vehicle batteries in battery swapping station. In *2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, pages 1–5, 2019.