2024-01-2006 Published 09 Apr 2024

Reinforcement Learning in Optimizing the Electric Vehicle Battery System Coupling with Driving Behaviors

Irem Altiner Temple University

Shiqi (Shawn) Ou South China University of Technology

Citation: Altiner, I. and Ou, Shiqi (Shawn) "Reinforcement Learning in Optimizing the Electric Vehicle Battery System Coupling with Driving Behaviors," SAE Technical Paper 2024-01-2006, 2024, doi:10.4271/2024-01-2006.

Received: 18 Oct 2023 Revised: 23 Jan 2024 Accepted: 24 Jan 2024

Abstract

Battery Run-down under the Electric Vehicle Operation (BREVO) model is a model that links the driver's travel pattern to physics-based battery degradation and powertrain energy consumption models. The model simulates the impacts of charging behavior, charging rate, driving patterns, and multiple energy management modules on battery capacity degradation. This study implements reinforcement learning (RL) to the simplified BREVO model to optimize drivers' decisions on charging such as charging rate, charging time, and charging capacity needed. This is done by a reward function that considers both the driver's daily travel demands and the minimization of battery degradation over a year. It shows that using appropriate charger type (No Charge, Level 1, Level 2, direct-current Fast Charge [DCFC], extreme Fast Charging [xFC]) with an appropriate charging time can reduce battery degradation and total charging cost at the end of the year while satisfying driver's daily travel demand. Using the Level 2 charging every day for night charging can reduce the battery capacity by 1.3819 ‰ whereas following the charger type and charging time suggestions of the RL will bring this number down to the level of 0.8037 ‰ over a one-year timespan. This gap between degradation rates gets bigger when one prefers using DC FC or xFC only respectively. Based on their daily travel demands, this RL model provides valuable strategic guidance to drivers to increase the battery lifetime and minimize the total cost of owning an electric vehicle.

Introduction

The U.S. has committed to reaching net-zero greenhouse gas (GHG) emissions no later than 2050 and has a goal of reducing GHG emissions by 50–52 % compared to 2005 levels by 2030 [1]. This objective has caused numerous automotive manufacturers to redirect their product strategies towards electric vehicles (EVs). To reach this goal and to progress to a cleaner transportation system, extensive research on the power source of EVs is required and crucial. This study uses reinforcement learning (RL) aiming at creating an artificial intelligent driven model for the vehicle system by integrating the consideration of driving behavior data; which can facilitate an understanding of how to efficiently and effectively move from the current petroleum-based transportation energy system to one that is more sustainable, intelligent and energy-diverse.

RL is one of the subfields of machine learning that consists of an agent trying to learn an optimal behavior by interacting with its environment through trial and error [2]. The agent receives a reward or penalty after each of its actions and tunes its behavior accordingly. The goal of the agent is to find the actions that maximize the cumulative reward. The machine learning approach is not new in the studies related to EVs. Some studies develop feature-based machine learning models for estimating the battery capacity of different battery types using real-life data and considering different driving conditions such as mileage and time of the day [3]; and some studies use these models to predict battery health using the chemistry and physics of batteries based on the environmental conditions such as temperature [4, 5]. In addition, in [6] and [7] the authors use several machine learning algorithms to predict the aging of the battery including K-Nearest Neighbors, Decision Trees, Support Vector Machines, and Deep Learning. Besides the battery aging estimation, some studies aim to reduce charging costs. In [8], authors employ Deep Learning to propose a cost-efficient charging strategy based on the environmental conditions, pricing, and driving data. Moreover, there are also studies using RL in the EV setting. More specifically, there exist studies that take advantage of RL to maximize battery lifetime by optimizing cabin thermal management and traction [9], or controlling the power flow distribution to minimize RMS battery current. [10] Many studies consider machine learning algorithms, more specifically RL algorithms to focus on charging cost [11, 12, 13] or focus on battery lifetime only [14, 15]; and barely a few studies consider combining both aspects including drivers' travel demands and patterns, or a variety of charger types.

The Battery Run-down under Electric Vehicle Operation (BREVO) is model is one of the models that links the driver's travel pattern to physics-based battery degradation and powertrain energy consumption models [16]. The model simulates the impacts of charging behavior, charging rate, driving patterns, and multiple energy management modules on battery capacity degradation. It offers a valuable assessment of how different chargers impact battery degradation, drawing from drivers' travel behaviors. Additionally, it provides an approximation of the comprehensive ownership costs associated with EVs. In alignment with the BREVO model, taking into account real-world statistical data regarding the daily travel patterns of drivers in the U.S., this study employs RL to offer drivers recommendations for both charger selection and optimal charging durations. The objective is to balance cost-effectiveness and battery lifetime. This aims to address the daily travel needs while also minimizing the total cost of charging and potential battery degradation resulting from the charging process. Choosing the best charger type every day for the battery lifetime may result in the least amount of battery degradation and cost after a year, however, it may cause not having enough battery charge for completing the day, especially for the days in which the travel demand is high. On the other hand, choosing a faster charger type every day may cause more battery degradation and a higher cost even though it ensures enough battery charge to satisfy the travel demands of every day. For these reasons, it is crucial to find a balance that optimizes battery degradation and charging costs. It must be noted that the BREVO model is one of the unique models that combine drivers' travel patterns and different charger types to compute the real-time energy consumption and the battery degradation rate. To do this, it considers variety of aspects such as air temperature, HVAC usage, braking regenerative system, etc. This study takes advantage of the BREVO model's original idea of combining drivers' travel patterns and different charger types and uses it to optimize battery degradation-charging costs balance while satisfying the drivers' travel demands. Instead of using the real-time energy consumption and battery degradation rates that BREVO outputs, this study uses fixed numbers described in the following sections as a start. The next steps of this research include incorporating the outputs of BREVO into the RL model.

This paper consists of four sections. Section one discusses some of the literature on the field and provides the motivation and objective of the study. Section two delivers a summary of the theory behind RL. Section three explains data, assumptions, and methodology used in this study. The last section provides summaries and conclusions.

Reinforcement Learning

RL involves learning what to do in an unknown environment or what actions to take based on the current state to maximize a numerical reward [2]. It is one of the three basic machine learning paradigms along with supervised learning and unsupervised learning.

There are four main elements of RL: an environment, an agent, a policy, and a reward. An agent is someone who learns and makes decisions. The outside world that the agent lives in and interacts with is called the environment. The process revolves around an active decision-maker, the agent, interacting with its environment, where the agent strives to achieve a goal despite not always knowing what to expect from the environment. The agent's decisions are called actions. The actions are allowed to impact the future state of the environment (e.g., the car's future battery capacity and the future battery charge level), therefore impacting the actions available to the agent at later times. Making the right decision involves considering the future outcomes of actions, which may, in turn, call for the ability to anticipate or plan ahead.

A policy, denoted by π, outlines how the learning agent should behave at a specific moment. In simpler terms, it is like a guide that tells the agent what actions, A_t, to take at a time t when it encounters certain states, S_t, in the environment.

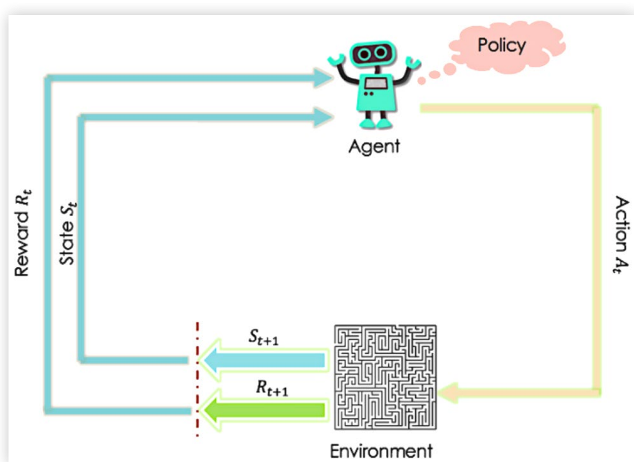
A reward at a time t, denoted by R_t, is a numerical value the environment sends to the agent as a result of the agent's decision/action and the current state of the environment.

Figure 1 diagrams the agent– environment interaction.

The agent's goal it to maximize the expected cumulative discounted reward:

$$G_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \ldots = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}, \quad \gamma \in [0,1]$$

FIGURE 1 The agent-environment interaction in RL



There exist different RL methods for finding the optimal solutions. The remaining part of this section focuses on the Q-Learning algorithm which is the algorithm used for the EV setup.

Q-Learning

Under a fixed given policy π, let $Q(S_t, A_t) := \mathbb{E}\left[G_t \middle| S_t, A_t\right]$ be the expected cumulative discounted reward after taking the action A_t in the state S_t, at a time t.

The target (or optimal) Q-value for this policy is $Q_*(S_t, A_t) := R_{t+1} + \gamma \max_{A \in \mathcal{A}(S_{t+1})} Q(S_{t+1}, A)$, as the goal is to take the action that will maximize the expected cumulative reward in the next state. Here, $\mathcal{A}(S_{t+1})$ is the space of actions available at state S_{t+1}.

The agent starts with an initial state S, chooses an action using the policy π derived from the Q-value, observes the reward R and the next state S', and the current Q-value of the state S is updated to

$$Q(S,A) \leftarrow Q(S,A) + \alpha.\left[Q_*(S,A) - Q(S,A)\right]$$
$$\leftarrow Q(S,A) + \alpha.\left[R + \gamma \max_A Q(S',A) - Q(S,A)\right]$$

where $\alpha \in [0,1]$ is called the learning rate and it determines how fast the Q-value is updated. The process continues until the agent reaches its final state.

Provided that all state-action pairs are visited infinitely many times, Q has been proved to converge with probability 1 to Q_*.

Method and Assumptions

As mentioned in the first section, the purpose of this study is to implement RL in an EV set up to provide guidance for drivers to maximize battery lifetime and minimize the total charging costs over a one-year timespan while meet drivers' daily travel demands.

The agent in the RL algorithm corresponds to the driver in this setting. The environment on the other hand includes information of the driver's daily travel demand, the EV's daily energy consumption, and available charging times at night. In every day of a whole year (365 days), the driver starts the day with a specific amount of charge in the battery and this battery charge decreases based on the mileage and the energy consumption of that day. The agent's task is to find the best charger type among five different options with the best charging time that will minimize the battery degradation and the cost of charging after a year while meeting the next day's travel demand by a driver. The decisions of charger types and charging times form the actions the agent or the driver chooses. Based on the action of charger type and the charging time, battery capacity and battery charge get updated, and this updated information builds the current state of the RL algorithm. This current state determines whether the driver's travel demand on the next day will be satisfied or not. Figure 2 shows the agent – environment interaction in this specific setting.

Data Used

Daily Travel Demand Based on the Bureau of Transportation Statistics, 2017, daily mileage information for 365 days is taken from a fixed gamma distribution with the shape parameter 1.92 and the scale parameter 15.20 [17]. Figure 3(a) shows the distribution of the data.

Daily Energy Consumption The daily energy consumption rate is taken from a uniform distribution with the lowest value of 13 kWh/100km and the highest value of 25.3 kWh/100km. The low and the high values are based on the general investigation results on EV energy

FIGURE 2 Agent-environment interaction in EV setting

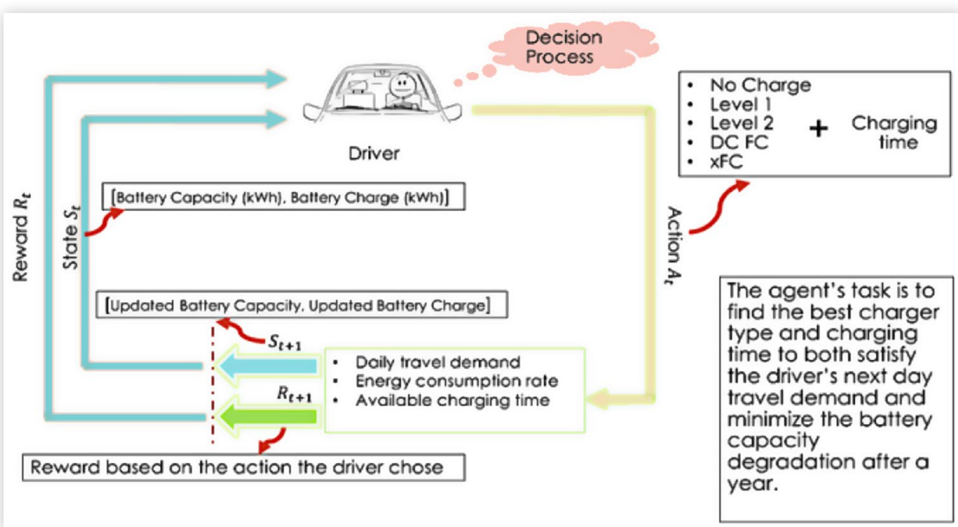
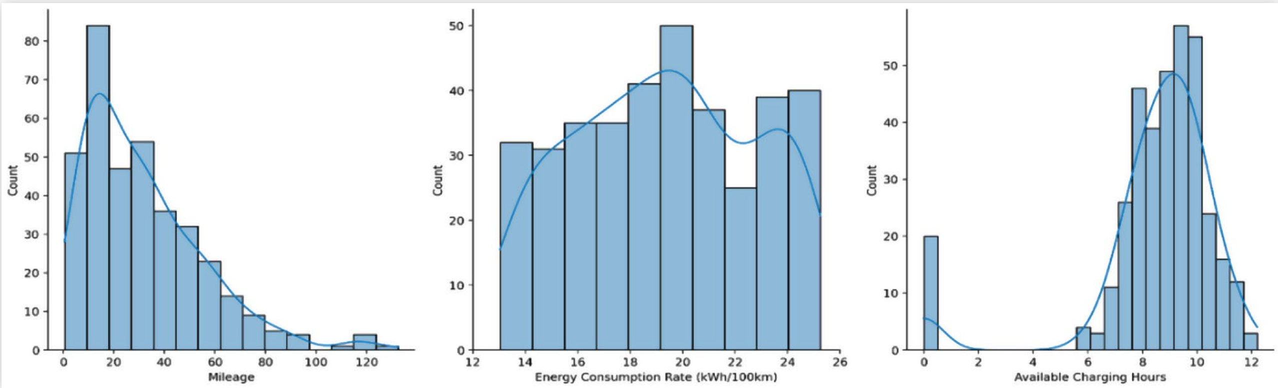


FIGURE 3 From left to right (a) Daily mileage; (b) daily energy consumption rate; and (c) daily available charging hours



consumption rates from the fueleconomy.gov [18]. For example, based on the comparison results in the fueleconomy.gov, the 2023 model Tesla Model S and Toyota bZ4X have an average of 17.4 kWh/100km daily energy consumption rate, whereas the 2023 Nissan Leaf SV has 19.3 kWh/100km and the 2023 Ford Mustang Mach-E AWD has 22.4 kWh/100km. This number is as low as 14.9 kWh/100km for the 2023 Hyundai Ioniq 6 Long range RWD 18-inch wheels [18]. For simplicity, this model assumes a uniform distribution of daily energy consumption rates between 13 and 25.3 kWh/100km. The distribution graph is given in Figure 3(b).

Available Charging Hours The available charging time is defined by the difference between the time that the driver arrives home in the evening and the time that they leave home for work/daily errands on the next day. This study assumes the available charging time is in the interval [0, 12] in hours and is taken from a fixed prescribed distribution, as shown in Figure 3(c). For simplification, the study assumes that the driver charges the EV only at home at night; and all charger types are optional for use.

Battery Capacity and Charger Types Based on 2024 Nissan Leaf SV Plus, the initial battery capacity of the vehicle is assumed to be 60 kWh [19]. In the model, four different charger type is considered. These are Level 1 charger, Level 2 charger, DC fast charger (DC FC) and extreme fast charger (xFC). The features of these charger types are shown in Table 1. Daily battery degradation rate gives how much of the battery becomes unusable after charging the vehicle with the selected charger type. For example, charging the vehicle with the Level 1 charger causes 0.002‰ of the initial battery capacity to be unusable; meaning that after using Level 1 charger on day one will decrease the battery capacity from 60 kWh to $60 - 60 \times \frac{0.0002}{100} = 59.9999$ kWh. Charging power determines amount of energy charged to the vehicle by time. For example, if the current charge of the battery is 40 kWh, then charging the battery with the Level 1 charger for 2 hours will bring the battery charge to $40 + 1.8 \times 2 = 43.6$ kWh; and the cost of charging will be $0.13 \times 1.8 \times 2 = \$0.468$. The battery model is from the BREVO model and described in [16] in detail.

Reinforcement Learning Elements

As stated above, the EV driver works as the agent in the algorithm. The remaining elements of the RL specific to this environment are explained in the following subsections.

Action Space The action space $\mathcal{A}$ is the space of all possible actions the agent can choose from. This model contains 70 different actions in the form of an array [ChargerType, ChargingTime], where the charger type could be one of the four charger types listed in the Battery Capacity and Charger Types section, plus the option of no charging; and charging time is any value in the set {0.25,0.5,1,2,3,...,12} in hours.

Observation Space Observation space $\mathcal{S}$ is the space of all possible states that the EV can have. At the end of each day, the agent/driver gets informed about the car's current state in the form of an array [BatteryCapacity, BatteryCharge]. The initial state is assumed to be $S_0 = [60, 60]$ in kWh. This space is continuous as both battery capacity and battery charge are values in the continuous interval [0, 60].

Reward After each action is calculated, if the driver chooses to charge, they receive a negative reward for causing battery degradation and increasing cost. The value of the rewards is determined by the rate of degradation and by the cost given in Table 1. If they choose no charge option, they receive a small positive reward.

The goal is to complete the year with wise charging decisions to have a good battery capacity with the least amount of cost. As the driver receives negative rewards if they choose to charge the vehicle, the cumulative reward mathematically decreases as days pass. Since the algorithm may stop before the time reaches the 365th day, and the objective is to maximize the cumulative reward, it is crucial to have more cumulative reward as

TABLE 1 Charging types and features used in the model

Charger Type	Charging Power (kWh)	Charging Price ($/kWh)	Daily Battery Degradation Rate
Level 1	1.8	0.13	0.002‰
Level 2	7.6	0.13	0.004‰
DC FC	60	0.26	0.006‰
xFC	400	0.52	0.01‰

more days proceeded. For this reason, after the appropriate rewards corresponding to the action, +70 is added to the reward, except for the very last step/action that causes the algorithm to stop. See the Algorithm 1 presented in Figure 4 for the reward function. In the pseudocode R_0 corresponds to the number portion of the daily degradation rate, R_1 corresponds to the charging power, and R_2 corresponds to the charging price from Table 1 corresponding to the chosen charging type.

Algorithm The driver starts Day 1 with the initial state

FIGURE 4 Reward Function

```
Algorithm 1 Reward Function
  function REWARD HANDLER(action)

      Reward ← 0
      if Charger Type ≠ No Charge then
          Reward ← −R0 − R1 · R2·(Charging Time)
      else
          Reward ← 0.001
      end if
      Reward ← Reward +70
      if Done and Day < 365 then
          Reward ← Reward −70
      end if
      return Reward
  end function
```

$$\left[\text{BatteryCapacity,BatteryCharge}\right] = S_0 = \left[60{,}60\right]$$

At the end of the day, based on the mileage information and the energy consumption rate of Day 1 the battery charge decreases. For example, if the energy consumption rate of the first day is 20kWh/100km (equivalent to 20kWh/62.137miles) and the driver traveled 30 miles, then the battery charge at the end of the first day becomes

$$60 - \frac{20}{62.137} \times 30 = 50.344\text{kWh}$$

The driver chooses an action of the form [ChargerType, ChargingTime]. For example [Level 2, 5] means that the driver charges their car with the Level 2 charger for 5 hours. As a result of this action, the battery capacity decreases, and the battery charge increases by the values in Table 1 and by the computations explained in the Battery Capacity and Charger Types section. Note that, the battery charge must be less than or equal to the battery capacity. The charging stops when the battery charge reaches its maximum, namely when it reaches the current battery capacity, regardless of how long the charging time has been chosen. The reward is received based on the updated time (if updated) the vehicle has been charged. At the end of this process, the new state becomes and the driver receives their reward according to the function presented in Figure 4.

$$S_1 = \left[\text{UpdatedBatteryCapacity, UpdatedBatteryCharge}\right],$$

Finally, whether the agent/driver will continue this cycle or not is determined by a function. If $S_1[0]$= UpdatedBatteryCapacity ≠0, and $S_1[1]$= UpdatedBatteryCharge is enough for next day's travel demand, then Day 2 starts, and the same steps are followed. If $S_1[0]$= UpdatedBatteryCapacity =0, or $S_1[1]$= UpdatedBatteryCharge is not enough for the next day's travel demand, then the cycle stops and the agent/driver starts over from Day 1. The Algorithm 2 given in Figure 5 shows the function that checks if the agent can continue its cycle the next day. Again, the agent's goal is to finish 365 days with the maximum reward.

Q-Learning and Hyperparameters As noted in the previous section, Q-Learning algorithm has been used for this setup. Q-Learning algorithm is an algorithm involves a Q-value for each state-action pair. It is noted that each state [BatteryCapacity, BatteryCharge] is an element of the continuous interval $[0,60] \times [0,60]$ meaning that there are infinitely many state that the agent can have. To finitize state-action pairs, the continuous interval [0, 60] has been discretized by dividing it into 60 pieces. As a result, the Q-value function is a function defined on a finite set of cardinality $|\mathcal{S}| \times |\mathcal{A}| = (60 \times 60) \times 70 = 252000$, which means there are 252000 Q-values for each possible discretized state-action pair.

FIGURE 5 Done Function

```
Algorithm 2 Done Function
  function TERMINATED(self)
      Done = False
      if Battery Capacity = 0 or Battery Charge < Needed Battery Charge
  then
          Done = True
      end if
      return Done
  end function
```

The initial Q-value for all discretized state S and action A is $Q(S,A) = 0$. The Q-values are updated after an action is taken by the rule explained in the RL section.

The agent/driver was trained in the environment using the algorithm explained in the previous section 60,000 times. This means the agent went through the 365-day cycle 60,000 times to figure out the best possible actions for each state. 60,000 is called the number of episodes. By the construction of the algorithm, not every episode lasted for 365 days. The total training time took no more than 30 minutes.

In the RL section, it was mentioned a policy and two hyperparameters, $\alpha \in [0,1]$, the learning parameter, and

$\gamma \in [0,1]$, the discount factor. These are the parameters to be chosen. In this study they are chosen as follows:

- Policy: The RL policy determines how the agent chooses its actions. It gives a ratio of exploration and exploitation. The exploration allows the agent to choose its actions randomly and possibly take different actions leading it to observe different states. The exploitation helps the agent choose the best action for the current state learned from the previous experiences.

  In this sense, the policy used in this work is an ε – greedy policy in which the actions are chosen randomly with probability ε, and the best action is chosen with probability $1 - \varepsilon$. In the beginning ε is set to $\varepsilon = 1$ and after each episode it is set to decrease by the exponential decay rate of 0.0001 with the minimum $\varepsilon = 0.01$. The formula of the decrease is as follows:

$$\begin{aligned} \varepsilon &\leftarrow \varepsilon_{min} + \left(\varepsilon_{max} - \varepsilon_{min}\right) e^{-decayrate \times episode} \\ &\leftarrow 0.01 + \left(1 - 0.01\right) e^{-0.0001 \times episode} \end{aligned}$$

  In this way the agent chooses its actions randomly in the beginning and explore more, and as the episodes pass by it starts choosing the best actions that it learned from the previous episodes with a larger probability to maximize the reward.

- Learning Parameter α: The beginning α is set to be $\alpha = 0.2$ and after each episode, it is set to decrease by the exponential decay rate of 0.00002 with the minimum $\alpha = 0.001$. A decreasing learning parameter is preferred because, in the beginning, the agent does not know much about the environment, so it is a good idea to update the q-values quickly and learn faster by taking big steps. As it gets more experienced in the environment it gets better at making decisions, so it means it is closer to its goal. In this stage, it is better to have a small learning rate for more precise and more stable q-value updates to avoid oscillations and getting stuck in the suboptimal solution. The formula of the decrease is as follows:

$$\begin{aligned} \alpha &\leftarrow \alpha_{min} + \left(\alpha_{max} - \alpha_{min}\right) e^{-decayrate \times episode} \\ &\leftarrow 0.001 + \left(0.2 - 0.001\right) e^{-0.00002 \times episode} \end{aligned}$$

- Discount Factor γ: Discount factor determines agent's preference for future rewards than the immediate rewards. A γ relatively close to 1 will cause agent to take actions considering the long-term effect of those actions, i.e., the future rewards more. This means, it is reasonable to decrease the discount factor, γ, at the end of every day (after every step/action taken) as the agent is not going to be around long enough to get long-term reward. So, it is better to focus on immediate rewards towards the end.

  In the beginning, γ is set to be $\gamma = 0.5$, and after each action taken (or after every day) it is set to decrease by the exponential decay rate 0.001 with the minimum $\gamma = 0.1$. Note that here the discount factor is decreased after each action instead of after each episode. After an episode is over, the discount factor is set back to its initial value of $\gamma = 0.5$. The formula of the decrease is as follows:

$$\begin{aligned} \gamma &\leftarrow \gamma_{min} + \left(\gamma_{max} - \gamma_{min}\right) e^{-decayrate \times episode} \\ &\leftarrow 0.1 + \left(0.5 - 0.1\right) e^{-0.001 \times episode} \end{aligned}$$

Results

To measure how well the model does, the results are compared to the results where the driver uses the same charger type every day of a year for the amount of hours they spend at home at night. Figure 7 presents the battery capacity decrease and cost increase of each case. The battery degradation chart in Figure 7(a) shows that using RL decisions to charge the vehicle causes the least amount of battery degradation compared to charging the vehicle with Level 2, DCFC, and xFC in each day of a year. The results show that using an xFC hurts 3.44 ‰ of the battery; using DCFC hurts 2.06 ‰ of the battery; and using the Level 2 charger hurts 1.38 ‰ of the battery if used every day during the whole time the driver is at home that night for a year. On the other hand, using the charger type and charger time decisions of the trained RL agent hurts only 0.8 ‰ of the battery after a year. How many times the RL agent chooses different charger types are presented in Figure 6. It appears that with the daily travel needs of the driver, charging the vehicle every night is not necessary.

Throughout the year, it is possible to satisfy the next days' travel needs by not charging the car or by charging it for a short time mostly with the Level 1 or the Level 2 charger. The RL agent chooses the No Charge option for 127 days; the Level 1 charger for 121 days; the Level 2

FIGURE 6 Charger Type Count of RL Decision

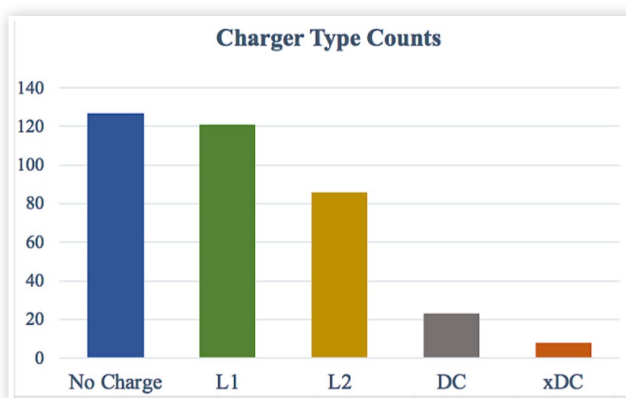
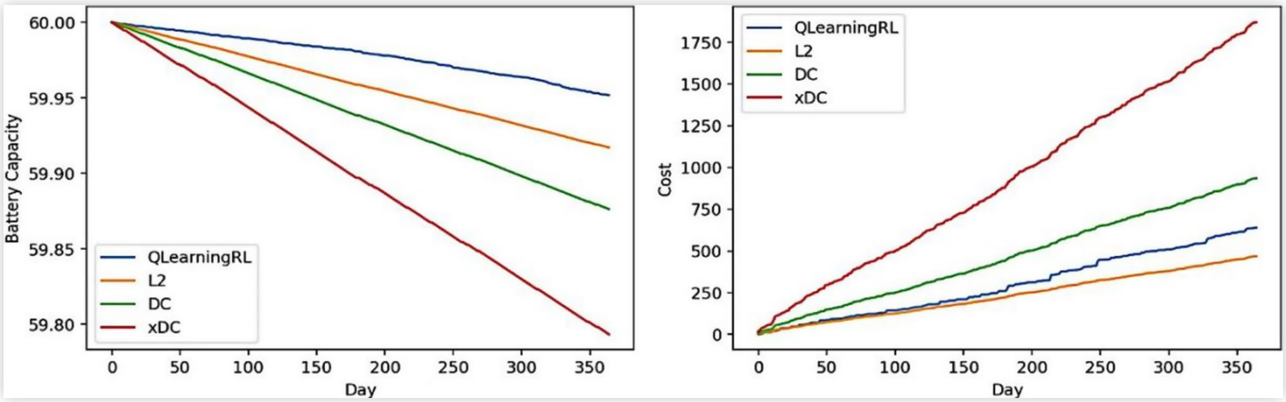


FIGURE 7 From left to right (a) Battery Capacity Chart; and (b) Cost Chart for Different Scenarios



charger for 86 days; the DCFC for 23 days; and the xFC for the remaining 8 days.

Note that, in the data used for the available charging time, there are 20 days that the available charging time is zero hours, meaning that the driver is not at home those days, so they do not charge their vehicle. In this case, the agent is forced to choose the No Charge option.

When the yearly cost chart in Figure 7(b) is examined, it comes into sight that it is more expensive to use the DCFC and xFC every day for a year on top of the fact that they cause more damage to the battery. On the other hand, it seems that using the Level 2 charger every day is slightly more budget-friendly than following the RL agent's suggestions. More precisely, the yearly cost of charging is $1868.63 with the xFC; $934.31 with the DCFC; $467.16 with the Level 2 charger; and $637.96 with the RL decisions. As seen from these numbers, using RL decisions costs around $170 more a year than using the Level 2 charger every day. However, when the current battery costs are considered, RL offers a better battery lifetime and total cost overall than using the same charger type every day.

According to the U.S. Department of Energy Fact of The Week #1272 published on January 9, 2023 [20], the cost of an EV lithium-ion battery pack declined 89 % between 2008 and 2022, however, the estimates for 2022 are still in the range of $153/kWh. This sums up to around $ 9180 for a 60-kWh battery pack. The fact that it provides results healthier for the battery, shows that RL can help consumers make wise charging decisions to prolong the lifetime of their vehicle and reduce the total cost of ownership.

The monthly costs of each scenario and the table containing the exact remaining battery capacities and total costs after a year are presented in Figure 8 and Table 2 respectively. In Figure 8, a cost comparison of each case is given and it can be deduced that RL can reduce charging costs to as low as $30 with an average of $53 per month, whereas these numbers are significantly higher for the DCFC and the xFC cases. As shown in Table 2, the exact capacity of the usable battery and the exact charging costs can be observed for each case.

FIGURE 8 The monthly cost of each scenario

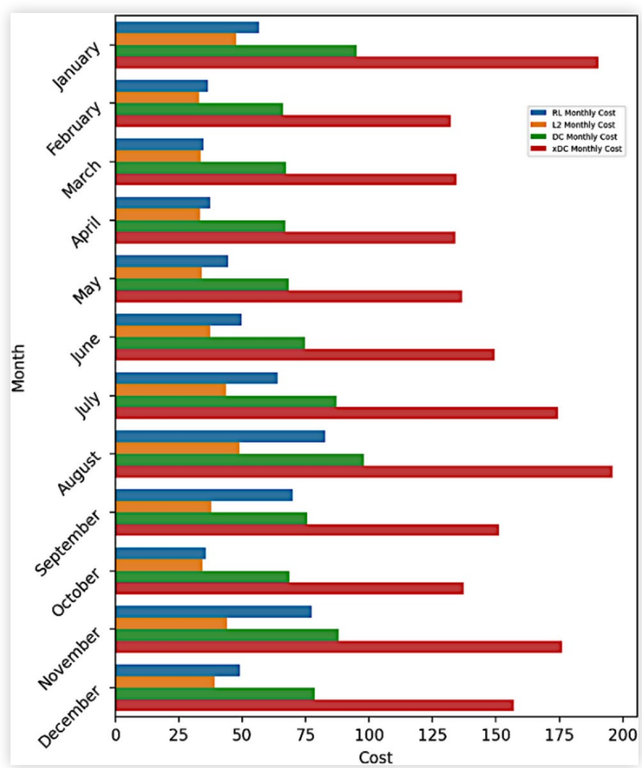


TABLE 2 Usable battery and cost table after a year

Charger Type	Usable Battery in the Beginning (kWh)	Usable Battery After 1 Year (kWh)	Total Cost After 1 Year
RL	60	59.95178	637.9561
Level 2	60	59.9171	467.1567
DCFC	60	59.8763	934.3134
xFC	60	59.7934	1868.6269

It can be seen that the figures do not include the case of using Level 1 charger every day. The reason for this is that considering the driver's daily available charging times and daily energy consumption, using a Level 1 charger every day that the driver is able to charge their vehicle will not be able to meet the driver's travel demands for more than one day. The driver must use a Level 2, DC fast, or an xFC to complete these days without charging during the day.

Conclusions

Following the existing BREVO model, this study employs RL into real-world statistics of drivers and vehicles in the United States. It demonstrates that given the daily mileage and energy consumption information of a driver, RL can suggest an appropriate charger type and charging time to offer a longer battery life and a better charging cost-saving. The study's contribution lies in employing RL considering both battery lifetime and charging cost aspects of total cost of ownership using the real-world statistics of drivers in the United States. However, there are a couple aspects of this work that should be noted.

1. The trained model is driver specific. The final policy depends heavily on the driver's daily travel demand and the available charging times the driver has. Because the agent decides whether to charge based on the available charging times and the next day's mileage information. This means the trained model may not perform well on another driver with different driving patterns.
2. Under the assumption that the driver will have similar driving habits each year, the trained model can be implemented for future operation with small modifications (such as the starting battery capacity should decrease). Otherwise, the model should continue to be trained with the fresh data from the driver.

Moreover, there are a couple aspects of this work that need further attention to improve.

1. Hyperparameters: Performance of machine learning algorithms closely related to the chosen hyperparameters. For this study, several of them have been tested and the best ones found were explained in the previous section. However, one can choose to run the algorithm with different hyperparameters to observe whether the results improve.
2. Data used: The study uses a data generated by several distributions with the parameters gathered from the Bureau of Transportation Statistics for the daily mileage data and previous studies for the daily energy consumption data. The next step is to incorporate the BREVO model into this study in more depth. BREVO model computes the daily mileage and energy consumption rate using real-life consumer time series data containing information on the speed of the vehicle at each second of the day, the temperature information of the cabin and the air, and several other factors such as HVAC system On/Off, etc. The next step of this study aims to run the algorithm discussed in this paper on the output data of the BREVO model to achieve similar results.
3. Constant charging cost and battery degradation rates: In real life, battery degradation rates and charging costs are not always constant. As a starting point, this study assumes that the battery degradation and the charging costs are constant. Unlike the BREVO model, the model also assumes that the battery degradation happens only by charging. However, under these assumptions this study achieves promising results and encourages a deeper work for more realistic scenarios. The future work of incorporating the outputs of BREVO model into the RL model will remove the assumptions of a constant battery degradation rate, and in addition, will enable the usage of real energy consumption rates instead of a fixed one.

Data Availability

The algorithm and model training files can be found on GitHub (https://github.com/altineri/EV-Environment)

Acknowledgments

This work conducted by I. A. was completed at Oak Ridge National Laboratory (ORNL) and Temple University, and supported in part by an appointment with the National Science Foundation (NSF) Mathematical Sciences Graduate Internship (MSGI) Program. This program is administered by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the U.S. Department of Energy (DOE) and NSF. The work conducted by S.O. and Z.L. was financially supported by the Pazhou Lab (PZL2023ZZ0009) and used the facilities from the South China University of Technology (SCUT). All opinions expressed in this paper are the authors' and do not necessarily reflect the policies and views of ORNL, Temple University, NSF, ORISE, SCUT, or Pazhou Lab.

Contact Information

Irem Altiner
irem.altiner@temple.edu
Phone number: 267-304-0925
Mailing address: Temple University, Wachman Hall, Department of Mathematics, 1805 North Broad Street, Philadelphia, PA 19122, USA

Shiqi(Shawn) Ou
sou@scut.edu.cn
Phone number: +86-020-81181684
Mailing address:
1. South China University of Technology, School of Future Technology, 777 Xingye Ave East, Panyu District, Guangzhou, Guangdong, 511442, China
2. Pazhou Lab, 70 Yuean Road, Haizhou District, Guangzhou, Guangdong 510335, China
oushiqi@pazhoulab.cn

References

1. "FACT SHEET: President Biden Sets 2030 Greenhouse Gas Pollution Reduction Target Aimed at Creating Good-Paying Union Jobs and Securing U.S. Leadership on Clean Energy Technologies," April 22, 2021, accessed: 6 October 2023, https://www.whitehouse.gov/briefing-room/statements-releases/2021/04/22/fact-sheet-president-biden-sets-2030-greenhouse-gas-pollution-reduction-target-aimed-at-creating-good-paying-union-jobs-and-securing-u-s-leadership-on-clean-energy-technologies/.
2. Sutton, R.S. and Barto, A.G., *Reinforcement Learning: An Introduction* (MIT Press, 1998)
3. Zhao, J., Ling, H., Liu, J., Wang, J. et al., "Machine Learning for Predicting Battery Capacity for Electric Vehicles," *eTransportation* 15 (2023).
4. Aykol, M., Gopal, C.B., Anapolsky, A., Herring, P.K., van Vlijmen, B., Berliner, M.D. «Perspective—Combining Physics and Machine Learning to Predict Battery Lifetime,» *Journal of The Electrochemical Society,* vol. 168, no. 3, 2021.
5. Jafari, S., Shahbazi, Z., and Byun, Y.C., "Lithium-Ion Battery Health Prediction on Hybrid Vehicles Using Machine Learning Approach," *Energies* 15, no. 13 (2022).
6. Harippriya, S., Vigneswaran, E.E., and Jayanthy, S., "Battery Management System to Estimate Battery Aging using Deep Learning and Machine Learning Algorithms," *Journal of Physics: Conference Series* 2325 (2022).
7. Kaushik Das, R.K.A.K., "Analyzing Electric Vehicle Battery Health Performance Using Supervised Machine Learning," *Renewable and Sustainable Energy Reviews* no. 189 (2024).
8. López, K.L., Gagné, C., and Gardner, M.A., "Demand-Side Management Using Deep Learning for Smart Charging of Electric Vehicles," *IEEE Transactions on Smart Grid* 10, no. 3 (2019): 2683-2691.
9. Haskara, I., Hegde, B., and Chang, C.F., "Reinforcement Learning Based EV Energy Management for Integrated Traction and Cabin Thermal Management Considering Battery Aging," *IFAC* 55, no. 24 (2022): 348-353.
10. Abdelhedi, R., Lahyani, A., Ammari, A.C., Sari, A. et al., "Reinforcement Learning-Based Power Sharing Between Batteries and Supercapacitors in Electric Vehicles," in *IEEE International Conference on Industrial Technology (ICIT)*, 2018.
11. Chang, F., Chen, T., Su, W., and Alsafasfeh, Q., "Control of Battery Charging Based on Reinforcement Learning and Long Short-Term Memory Networks," *Computers & Electrical Engineering* 85 (2020).
12. Hepeng Li, Z.W.H.H., "Constrained EV Charging Scheduling Based on Safe Deep Reinforcement Learning," *IEEE Transactions on Smart Grid* 11, no. 3 (2019): 2427-2439.
13. Valogianni, K., Ketter, W., and Collins, J., *Smart Charging of Electric Vehicles Using Reinforcement Learning*, 2013 .
14. Fei, Z., Yang, F., Tsui, K.L., Li, L. et al., "Early Prediction of Battery Lifetime via a Machine Learning Based Framework," *Energy* 225 (2021).
15. Yang, H., Hong, J., Liang, F., and Xu, X., "Machine Learning-Based State of Health Prediction for Battery Systems in Real-World Electric Vehicles," *Journal of Energy Storage* 66 (2023).
16. Ou, S., "Estimate Long-term Impact on Battery Degradation by Considering Electric Vehicle Real World End-use Factors," *Journal of Power Sources* 573 (2023).
17. "National Household Travel Survey Daily Travel Quick Facts," May 31, 2017, accessed July 2023, https://www.bts.gov/statistical-products/surveys/national-household-travel-survey-daily-travel-quick-facts.
18. "www.fueleconomy.gov the official U.S. Government Source for Fuel Economy Information," accessed October 2023, www.fueleconomy.gov.
19. "2024 Nissan LEAF® Range & Battery," 2023, accessed July 2023, https://www.nissanusa.com/vehicles/electric-cars/leaf/features/range-charging-battery.html.
20. "FOTW #1272, January 9, 2023: Electric Vehicle Battery Pack Costs in 2022 Are Nearly 90% Lower than in 2008, According to DOE Estimates," January 9, 2023, accessed October 6, 2023, https://www.energy.gov/eere/vehicles/articles/fotw-1272-january-9-2023-electric-vehicle-battery-pack-costs-2022-are-nearly.

ISSN 0148-7191