Optimizing Home Energy Management and Electric Vehicle Charging with Reinforcement Learning

Di Wu
McGill Uniersity
Montreal, Quebec
di.wu5@mail.mcgill.ca

Guillaume Rabusseau
McGill Uniersity
Montreal, Quebec
guillaume.rabusseau@mail.mcgill.ca

Vincent François-lavet
McGill Uniersity
Montreal, Quebec
vincent.francois-lavet@mcgill.ca

Doina Precup
McGill Uniersity
Montreal, Quebec
dprecup@cs.mcgill.ca

Benoit Boulet
McGill Uniersity
Montreal, Quebec
benoit.boulet@mcgill.ca

ABSTRACT

Smart grids are advancing the management efficiency and security of power grids with the integration of energy storage, distributed controllers, and advanced meters. In particular, with the increasing prevalence of residential automation devices and distributed renewable energy generation, residential energy management is now drawing more attention. Meanwhile, the increasing adoption of electric vehicle (EV) brings more challenges and opportunities for smart residential energy management. This paper formalizes energy management for the residential home with EV charging as a Markov Decision Process and proposes reinforcement learning (RL) based control algorithms to address it. The objective of the proposed algorithms is to minimize the long-term operating cost. We further use a recurrent neural network (RNN) to model the electricity demand as a preprocessing step. Both the RNN prediction and latent representations are used as additional state features for the RL based control algorithms. Experiments on real-world data show that the proposed algorithms can significantly reduce the operating cost and peak power consumption compared to baseline control algorithms.

KEYWORDS

Home Energy Management; Electric Vehicle Charging; Reinforcement Learning

1 INTRODUCTION

With the advancement of technology and increasing attention to environment protection, more and more electricity-driven products are introduced into people's daily life. Electric vehicles (EVs) are more efficient than traditional internal combustion engine vehicles [9] and they are adopted by more and more customers. However, the introduction of EVs brings a high power consumption burden on the power system and may even jeopardize the infrastructure of power grids if there is no proper energy management [26]. In the meantime, renewable energy generation is increasingly adopted for residential homes. Renewable energy generation such as wind and solar energy can provide home owners with cheap and clean energy, but it is quite intermittent and sensitive to weather conditions.

Smart grid using distributed renewable generation, advanced meters, energy storage, communication and computation tools can cope with these challenges. Energy management is a core issue for the smart grid [1] and can be beneficial for both the customers and utility companies. Recently, with the development of home based energy storage and controllers, energy management for the residential sector has attracted more and more attention [13]. However, energy management for residential homes is a difficult problem. The main challenges come from uncertainties on both the power-supply and power-demand sides. Moreover, we usually only have limited historical data for residential houses which makes energy management an even more difficult task.

In this paper, we consider a general case of home energy management with EV charging where battery specifications and limited amount of historical data (for the consumption and production) are available. We show how this problem can be formulated as a Markov Decision Process (MDP) that we propose to tackle using two model-free reinforcement learning algorithms: Neural Fitted Q Iteration (NFQ) and Deep Q-Network (DQN). The limited amount and heterogeneity of the available data makes implementing these algorithms challenging. We first show how a home simulator (**RLEnergy**) can be built from the available historical data and battery specifications, allowing us to enable the interactions with the RL algorithms needed for efficient exploration. The second challenge resides in how to incorporate the time series of base load power consumption in the state features of the MDP. Indeed, while only considering the power demand at the current time step may not be enough, incorporating too long a history into the state features could lead to overfitting [6].

To address this issue, we propose to model the base load power consumption using a recurrent neural network (RNN) and to enrich the state representation with both the RNN prediction for the next time step and its latent representation. The performance of the two algorithms are showcased on real-world data to optimize operating cost and peak power consumption where they show significant improvements compared to previous rule based and batch RL methods.

Related work. In [27], the authors propose an approximate model for EV arrival and present a building energy management control algorithm with this model. In [17], the authors present a nonlinear model for a building cooling system and then build a control algorithm over this nonlinear model. The performance of these proposed control algorithms highly depends on the accuracy of system dynamics modeling which may be not realistic for residential homes where accurate system dynamics are usually unavailable. As EV batteries can be seen as distributed energy storage, EV charging scheduling has attracted significant attention, a survey on recent EV charging

control strategies is presented [25]. In [14], the authors showed that with proper management, EV batteries can help stabilize the power grid and support large scale renewable energy adoption.

Reinforcement learning based control algorithms for smart grid are discussed in some recent papers. In [7], the authors propose a DQN based control strategy for storage devices in a microgrid. In [2], the authors propose to use Fitted Q-Iteration (FQI) to deal with smart home energy management. In [3, 24], total power consumption of EV charging fleet is learned by a batch reinforcement learning (BRL) algorithm and single EV charging is then scheduled with linear programming. To the best of our knowledge, there is no previous work studying the home energy management system (EMS) integrated with EV charging in one reinforcement learning framework.

Deep reinforcement learning (Deep RL) algorithms exhibit strong generalization capabilities in problems with complex state space. They have for example shown successful applications in problems with very large number of states such as playing Atari, Go games and other complex control tasks [15, 18, 19]. With the development of distributed monitors and controllers, smart homes control tasks have very complex state spaces. Deep RL algorithms, for their generalization abilities and strong representation power, could be promising candidates for home energy management where a large amount of features can be used. In this paper, we aim to investigate the performance of Deep RL algorithm (DQN) and NFQ on home energy management integrated with EV charging.

Contributions and outline. In this paper, we propose two RL based control algorithms to handle both the interactions with the power grid and EV charging scheduling in one unified RL framework. Our main **contributions** can be summarized as follows: 1) We propose an approach that can model smart home energy management with EV charging as a Markov decision process (MDP) [22]. 2) We tackle it with two model-free reinforcement learning algorithms: NFQ, DQN, and we investigate their performance on reducing operating cost and peak power consumption with real-world data.

After introducing relevant technical background in Section 2, the main components of smart home systems are introduced in Section 3. In Section 4, we show that energy management in smart homes can be formalized as an MDP and we propose two model-free RL based control algorithms to address it. Section 5 presents the experimental results with houses where only some historical data and battery specifications are available. Finally, Section 6 concludes the paper.

2 BACKGROUND

2.1 Markov Decision Process

Home EMS can be seen as a sequential decision problem and can be modeled using a Markov Decision Process. An MDP is a tuple $\langle S, A, T, R, \gamma \rangle$, where:

- S is a finite set of the states;
- A is a finite set of possible actions;
- $T : S \times A \times S \rightarrow [0, 1]$ is the transition probability from state s_t to state s_{t+1} when an action a_t is taken.
- $R : S \times A \rightarrow \mathbb{R}$ is the reward function, i.e. $R(s, a)$ is the reward received by the agent when taking action a in state s.
- γ is the discount factor $\gamma \in [0, 1)$

Solving an MDP means finding an optimal policy $\pi : S \rightarrow A$ maximizing the long-term cumulative reward G_t as shown in Equation 1.

$$G_t = \mathbb{E}[\sum_{j=1}^{\infty} \gamma^{j-1} R_{t+j}] \quad (1)$$

2.2 Batch Reinforcement Learning

Reinforcement learning can be used to solve an MDP when little or no knowledge of the system dynamics is available. In batch reinforcement learning (BRL), the data collection and the learning process are decoupled and the control policy is learned from a set of learning experiences built from a set of historical data [5]. The main objective of BRL is to learn the best control policy given the existing batch of learning experiences.

Neural Fitted Q Iteration [23] is one of the most popular BRL algorithms. NFQ converts the learning from interactions paradigm to a series of supervised learning processes. There are mainly three phases for NFQ: exploration phase, training phase, and execution phase. In the exploration phase, a batch of transition samples $F = \{(s_t, a_t, r_t, s_{t+1}) | t = 1, \ldots, T\}$ are gathered. In the training phase, a training set D_t^h is built with F iteratively: it associates tuples (s, a) with estimated Q values $\overline{q}_{s,a}^h$. For each iteration h, the Q value $(\overline{q}_{s,a}^h)$ for state action pair (s, a) is updated. A neural network is used to approximate the Q value function on D_t^h. In the execution phase, the policy learned in the training phase is applied.

2.3 Deep Q Network

Reinforcement learning is known to be unstable when a nonlinear function such as neural networks is used as function approximator. There are several reasons for this: the sequence of observations for reinforcement learning are correlated and the data distribution will change during the learning process. This violates the assumption that the data should be independently and identically distributed. NFQ tackles instability issues by learning the function approximator using hundreds of iterations. However this method is inefficient for large neural networks. Another potential reason for instability is the extrapolation abilities of neural networks. The contraction mapping property of the Bellman operator is not enough to guarantee convergence and one can have instabilities [8].

Deep Q network (DQN) proposed in [18, 19], has successfully overcome the aforementioned challenges on combing deep learning with reinforcement learning framework. DQN uses a deep neural network to approximate the Q-value function and uses two techniques to tackle instability issues. It uses experience replay [16] and a target network to reduce the correlations in the sequence observations and smooth the data distribution changes. In DQN, the Q value function $Q(s, a; \theta)$ is approximated by a deep neural network with parameters θ. In the learning process, the Q-learning updates are based on a mini-batch of experiences (s_t, a_t, r_t, s_{t+1}). This mini-batch is drawn uniformly at random from the pool of stored transition samples.

3 SMART HOME COMPONENTS

Figure 1 shows five main components of the smart home discussed in this paper. These components are: home energy management system, renewable energy generation, power grid, home based battery storage, and power demand including base load power consumption

and EV charging power consumption. Advanced meters are assumed to be installed in the home, enabling bi-directional communication.

3.1 Base Load Power Consumption

More and more electronic appliances are introduced in people's daily life. The power consumption of the smart home is mainly divided into two groups: EV charging power consumption and base load power consumption. We consider the base load power consumption including all other power consumption in the home except the EV charging, namely power consumption for home appliances, cooling, heating, fans, interior lights, etc. For base load power consumptions, there is a morning peak after inhabitants wake up and an evening peak when they arrive home.

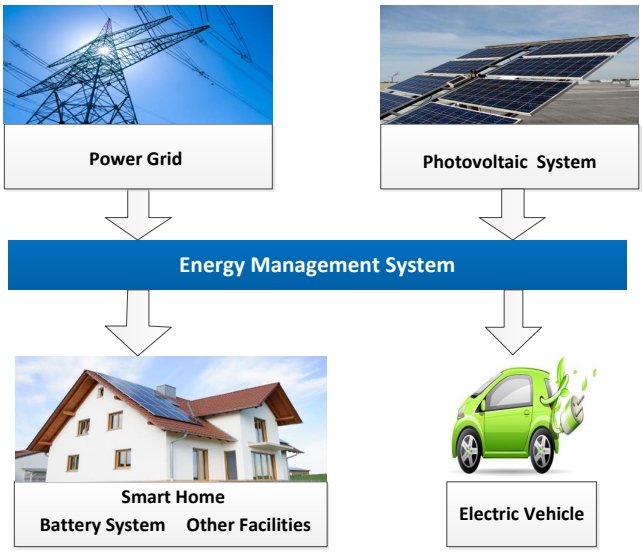

Figure 1: Energy management system components for smart grid

3.2 Electric Vehicle

EV charging has become one of the major power demand for the residential sector with the fast increase of EV adoption. Here, we consider that each residential home is integrated with one EV and the EV is only charged at home. As shown in [28], EVs usually leave home around 7 am and come back around 6 pm. We apply the constraint that the EV needs to be charged with enough energy before 7am. Meanwhile, we assume that the EV can be charged with continuous rate which means any power rate ranging from zero to the maximal allowed charging rate. Figure 2 shows the total power consumption for one home with a Honda Fit EV [11]. We can see that without any management, EV charging would coincide with the peak hours of base load power consumption and further increase the peak of total power consumption. This effect could be even worse for some residential networks, where early EV adopters may exhibit similar customer behaviors in the communities and appear in clusters [12].

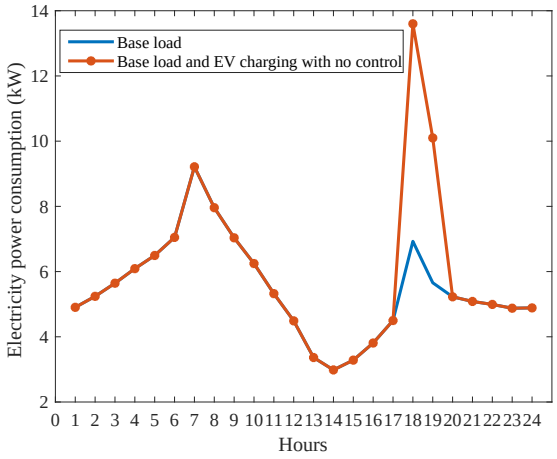

Figure 2: Electricity demand with EV charging

3.3 Photovoltaic Output

Home based renewable energy generation including solar and wind power generation are becoming an important power source for more and more homes. In this paper, we assume that the homes are equipped with solar panels which enable solar power generation. However, the output of solar power can be quite intermittent and varies according to different kinds of weather conditions.

3.4 Battery System

Home energy storage is an important component for smart homes. In this paper, we assume that a home based battery system is installed. The battery system can be used to save the energy when there is power surplus for later use and mitigate the volatility of renewable energy generation. We assume that home battery can be charged and discharged with continuous rate which means any power rate ranging from zero to the maximal allowed charging rate.

3.5 Home Energy Management System

For each time step, home EMS will make control decisions. Between two control time steps, the control actions will remain the same as the latest ones.

4 REINFORCEMENT LEARNING FOR HOME ENERGY MANAGEMENT

In a home EMS, at every time step (each hour in this paper) the decision maker has to interact with the power grid, the home battery charging scheduling and the EV charging scheduling. We start by showing how this sequential decision making problem can be formalized as an MDP and then propose two model-free reinforcement learning based control algorithms to solve it. In order to consider a more general scenario, we assume that we only have access to historical data of electricity demand, solar power generation, electricity price, EV arrival time, home based battery state of charge, EV departure time and that battery specifications are available.

4.1 Energy Management as an MDP

4.1.1 States. Optimal control actions for home EMS is determined by observing the current system state. Such a state (s_t) is composed of the following observations : home based battery state of charge (BSOC) ($H_{s,t}$), home based solar power generation ($H_{g,t}$), electricity price (P_t), base load power consumption ($H_{b,t}$), EV charging availability ($E_{a,t}$), time to departure for the EV ($E_{d,t}$), and current EV battery state of charge ($E_{s,t}$). $H_{s,t}$ is defined as 100% when the battery is fully charged and 0% when fully discharged. The last three state variables are EV related variable:. $E_{a,t}$ shows the charging availability for EV (set as 1 when EV is at home and 0 otherwise), $E_{s,t}$ shows the battery state of charge of the EV battery, and $E_{d,t}$ shows how many hours are left before the departure of the EV. Then the MDP state at time step t, $s_t \in S$ is defined as: $s_t = (H_{s,t}, H_{g,t}, H_{b,t}, P_t, E_{a,t}, E_{d,t}, E_{s,t})$.

It is worth mentioning that this MDP formulation is only an approximation of the home EMS since it is unrealistic to assume that the time series variables such as P_t and $H_{b,t}$ are Markovian. In Section 4.2, we will propose to enrich the state features using latent representations learned by an RNN from the time series, which potentially addresses this concern since such latent representations could intuitively capture the non-Markovian dynamics of the environment in the model (by encoding the relevant information from the time series).

4.1.2 Actions. In this paper, we implement control for the charging scheduling of EV batteries and home EMS interactions with power grid. We assume that both EV battery and home battery can be charged or discharged with continuous values (from zero to the maximal allowed charging rate) and home EMS can inject energy back into the power grid. For every time step, home EMS need to decide actions for EV charing $C_{e,t}$ and interactions with power grid $U_{b,t}$. Positive values of $U_{b,t}$ corresponds to power bought from the grid and negative values to power sold back to the power grid. $C_{e,t}$ corresponds the EV charging rate, where negative values correspond to discharging the EV with $C_{e,t}$. The power balance shown in Equation 2 must be satisfied for every time step (the left hand side is the power demand and the right hand side is the power supply). $C_{h,t}$ is the charging rate for the home battery and is decided deterministically as a function of $U_{b,t}$ and $C_{e,t}$. In this paper, we use four discretizations for both actions $U_{b,t}$ and $C_{e,t}$, leading to 16 possible actions at every time step.

$$C_{e,t} + C_{h,t} + H_{b,t} = H_{g,t} + U_{b,t} \tag{2}$$

After these actions are taken, the state features corresponding to home and EV BSOC are updated using Equation 3 and 4, where $H_{s,t+1}$ and $E_{s,t+1}$ are the SOC at time $t+1$ for home battery and EV battery respectively, B_h is the battery capacity for home battery, B_e is the EV battery capacity, and η is charging efficiency. In this paper, we assume $\eta = 0.9$.

$$H_{s,t+1} = H_{s,t} + \frac{C_{h,t}}{B_h}\eta \tag{3}$$

$$E_{s,t+1} = E_{s,t} + \frac{C_{e,t}}{B_e}\eta \tag{4}$$

At each time step, we first check whether the EV is in the home. If EV is not at home, all EV related variables are set to 0. If EV is at home, we need to determine the EV charging or discharging rate $C_{e,t}$. Then with the EV charging rate, renewable energy generation and base load power demand we can determine whether there is any power surplus: $C_t = H_{g,t} + H_{s,t} \cdot B_h - H_{b,t} - C_{e,t}$. If C_t is negative, then we need to buy energy from the power grid, if it is positive, we can sell energy back to the power grid. When deciding the allowed actions, the constraints of battery charging and discharging limits should always be satisfied which means that the charging or discharging rate should not more than rated maximal charging power.

4.1.3 Reward. The objective for home EMS is to reduce the long-term operating cost. We use negative cost as shown in Equation 5 as the MDP reward.

$$R_t = -Cost_t = -U_{b,t} * P_t \tag{5}$$

While in practice the buying and selling prices could be different for certain utility programs, we assume here that they are the same for the sake of simplicity.

4.2 Enriching the State Features with Recurrent Neural Networks

RL based control algorithms make control decisions based on current observations. It is intuitively clear that if future electricity demand is known, we could get better control policies. In this paper, we use long short-term memory (LSTM) [10] recurrent neural network (RNN) to model short-term electricity load forecasting. After training an LSTM network with k hidden units to predict the future base load power demand from historical power consumptions, we can thus enrich the state features of the MDP at each time step t with the demand prediction. Moreover, we include the latent representation of the LSTM a this time step which potentially encode relevant information on the trend of the demand time series. The MDP state with additional state features is now shown in Equation 6.

$$s(t) = (H_{s,t}, H_{g,t}, H_{b,t}, P_t, E_{a,t}, E_{d,t}, E_{s,t}, \hat{P}_{t+1}, \psi_L) \tag{6}$$

where $\hat{P}_{t+1}$ is the predicted based load power consumption for next time step and $\psi_L \in \mathbb{R}^k$ is the latent LSTM representation for time $t+1$. In this paper, we use time-of-use (TOU) electricity price which is given by the utility company and fixed for certain hours of the day. Thus, we do not implement prediction for electricity price.

4.3 Neural Fitted Q Iteration based Home Energy Management

Neural fitted Q iteration is an instance of the Fitted Q Iteration family [5] which uses a neural network to approximate the Q value function. NFQ allows the RL agent to learn a control policy from historical data. The NFQ based EMS control algorithm (NFQEMS) is defined in Algorithm 1.

The first phase for NFQ is exploration. To generate the transition experiences, we first capture historical data including solar energy generation, base load power consumption, EV arrival time, EV arrival BSOC, and EV departure time. We then assign a random state to the home battery for the first time step and take random actions from the allowed action sets for all following time steps, from which

Algorithm 1 NFQ based Home Energy Management System: NFQEMS

Require: Load $F = \{(s_t, a_t, r_t, s_{t+1}) | t = 1, ..., T\}$
Require: Define $\overline{Q}^0(s, a) = 0, \forall (s, a) \in F$, and $\overline{q}^h_{s,a} \in \overline{Q}^h(s, a)$
Require: Define H as the Horizon to be performed
Require: Define D^0_t as an initially empty training set
h = 1;
while $h \leq H$ **do**
  for all $(s_t, a_t, r_t, s_{t+1}) \in F$ **do**
    $\overline{q}^h_{s,a} = r_t + \gamma \max_{a \in A_{s(t+1)}} \overline{Q}^{h-1}(s_{t+1}, a)$;
    if $(((s_t, a_t)),) \in D^{h-1}_t$ **then**
      $D^h_t \leftarrow D^{h-1}_t - \{((s_t, a_t), .)\}$;
    end if
    $D^h_t \leftarrow D^{h-1}_t \bigcup \{((s_t, a_t), \overline{q}^h_{s,a})\}$
  end for
  Implement supervised learning
  Use supervised learning to train a function approximator $\overline{Q}^h_{s,a}$ on the training set D^h_t
  $h \leftarrow h + 1$
end while
Use the learned policy for smart residential home energy management

we get a set of transition experiences $F = \{(s_t, a_t, r_t, s_{t+1})\}$. In the training phase, NFQEMS learns an approximator for the Q value function using a training set $D_t = \{(s_t, a_t, q_t)\}_t$ built from the transition experience set F in the following way.

We first assign 0 to Q values for all state action pairs. We then learn an approximator with this initial training set D^0_t. In the following iterations of NFQEMS, we update the training set with new Q values as shown in Equation 7. The training phase continues until either the maximum number of iterations H or a convergence criterion is reached. In this paper, we use a feedforward neural network as the function approximator[1]. For NFQEMS, we set $H = 200$ and the convergence criteria is achieved when the variation of Q value is less than 5%, i.e. when the mean average percentage error for Q values of all state-action pairs of recent two iteration is less than 5%. In the execution phase, we can use a greedy policy with the learned Q value function approximator. The RL agent will choose an action from the allowed action set which has the highest Q value as follows, where γ is the discount factor.

$$\overline{q}^h_{s,a} = r_t + \gamma \max_{a \in A_{s(t+1)}} \overline{Q}^{h-1}(s_{t+1}, a) \tag{7}$$

4.4 Deep Q Networks based Home Energy Management

To use DQN for home energy management, we need a home simulator that the on-line RL algorithm can interact with. In this section, we first show how to build a smart home simulator, **RLEnergy**, from given historical data and battery specifications and then use this simulator to interact with DQN.

[1]We use a feedforward neural network with three hidden layers each with 128 neurons to approximate the Q function. Rectifier linear units (ReLU) is used as activation functions.

Algorithm 2 DQN based Home Energy Management System: DQNEMS

Initialize replay memory $M = [empty\ set]$ as Capacity N
Initialize neural network Q with random parameters θ
Initialize target network $\hat{Q}$ with parameters $\theta^- = \theta$
for $episode = 1, K$ **do**
  Go to time step $t = 1$ with RLEnergy simulator and assign a random value for home battery
  for t = 1, T **do**
    With probability ϵ select a random action
    Otherwise select $a_t = \underset{a}{argmax} Q(s_t, a_t; \theta)$
    Execute action a_t for home simulator and observe reward r_t and next state s_{t+1}
    Store transition (s_t, a_t, r_t, s_{t+1}) in replay memory M
    Sample random minibatch D_m of transitions (s_t, a_t, r_t, s_{t+1}) from M
    Set $y_t = \begin{cases} r_t & if\ \ t = T - 1 \\ r_t + \gamma \underset{a_{t+1}}{max} \hat{Q}(s_{t+1}, a_{t+1}; \theta^-_i) & otherwise \end{cases}$
    Perform a gradient decent step on $(y_t - Q(s_t, a_t; \theta))^2$ with respect to network parameters θ
    Update the target network parameters θ^- every C steps: $\theta^- = \theta$
  end for
end for
Use the learned policy for smart residential home energy management

RLEnergy. Suppose we have a historical data set D_h for T time steps. As defined in Section 4.1.1, for time step t, the state s_t is defined as $s_t = (H_{s,t}, H_{g,t}, P_t, H_{b,t}, E_{a,t}, E_{d,t}, E_{s,t})$. There are two kinds of state variables for s_t: fixed state variables and adaptive state variables. Within D_h, the state variables $H_{g,t}$, P_t, $H_{b,t}$, $E_{a,t}$ and $E_{d,t}$ are fixed for all time steps; $H_{s,t}$, $E_{s,t}$ will be updated according to the chosen actions $C_{h,t}$ and $C_{e,t}$ using to Equations 3 and 4. We can then build a simulator based on the historical data D_h and battery dynamics. For time step $t + 1$, s_{t+1}, $H_{g,t+1}$, P_{t+1}, $H_{b,t+1}$, $E_{a,t+1}$, $E_{d,t+1}$ are taken from D_h directly according to time step index and $H_{s,t+1}$ and $E_{s,t+1}$ will be decided on the actions of $C_{h,t}$ and $C_{e,t}$ taken at time step t. This simulator will enable us to use on-line reinforcement learning algorithms such as DQN.

DQN based home energy control algorithm DQNEMS is defined in Algorithm 2. We first initialize a replay memory M with capacity N. For DQNEMS, we parameterize the Q-network with parameters θ and target neural network $\hat{Q}$ with θ^-. We first assign $\theta^- = \theta$. As shown in Algorithm 2, the outer loop learns DQN with K episodes, and the inner loop shows the parameter updating for every time step. T is the total number of time steps for the used historical data. Every episode will end when T is reached.

For every time step, RL agent will take a random action with probability ϵ for exploration or choose the action with maximal Q value estimate. After action a_t is taken, immediate reward r_t is received and the agent will go to next state s_{t+1}. Transition tuple (s_t, a_t, r_t, s_{t+1}) will then be stored to memory M. An mini-batch of transitions D_m will be sampled randomly from M. Parameters

Table 1: Specifications for home battery and EV

Method	Capacity (kWh)	Maximal charging rate (kW)
Home Battery	10	1
EV [11]	20	6.6

for target network θ^- will be updated every C steps to stabilize the learning process. After K episodes, we can use the learned model for home energy management. [2]

5 EXPERIMENTAL RESULTS

5.1 Experiment Setup

In this section, we compare the performance of the proposed control algorithms with two baselines: rule-based and batch RL based RLbEMS which are described in [2]. Rule-based control will charge the EVs when they arrive and sell the energy to the grid when there is power surplus. The main differences between RLbEMS and our NFQEMS is that we consider the modeling and charging scheduling of EVs while RLbEMS only consider interactions between home and power grid. EVs will be charged when they arrive.

Time-of-use electricity price [4] and residential home power consumption for three houses in three locations in New York [21] are used. We have hourly electricity price and base load power consumption for one year. Figure 3 shows load consumption on a winter day for one house and electricity price structure. We can see that there are two peak power consumption periods: 6 - 9am and 6-10pm. For TOU electricity price structure, the peak hour is from 8am to 8pm. We can see that during the peak hours, electricity price is quite higher than that in the off-peak hours. We consider that there are 10 PV panels (peak power generation is 220W for each pannel) for every house and use the irradiance data from [20] to generate the solar power generation for one year. In this paper, we assume that every house has one EV (Honda Fit) and one home based battery. The specifications for EV and home battery are shown in Table 1. EV usage data described in [28] is used to build EV arrival data for one year. We assume that both the home battery and EV battery can be charged and discharged with continuous values from zero to maximal allowed charging rate. We use the first 11 months as a training set while the remaining data for one winter month is used as a testing set.

5.2 Operating Cost and Peak Power Reduction

The total operating cost for the test set for three houses under different control algorithms are shown in Table 2. We can see that the two proposed RL based control algorithms can help reduce operating cost compared with two baselines. For NFQEMS it can reduce by 6.71% of the operating cost over rule-based control and by 3.93% over RLbEMS control. For DQNEMS it can reduce by 6.91% of the operating cost over rule-based control and by 4.12% over RLbEMS control.

Table 3 shows that both NFQEMS and DQNEMS can help reduce average of daily operating cost for all houses. However, the standard deviation of average daily operating cost is quite high. This is probably due to the fact that the base load power consumption varies from day to day.

[2] To implement a fair comparison with NFQ, we use the same neural network structure as described in Section 4.3, except that the output layer has 16 neurons corresponding to the Q value for every possible action.

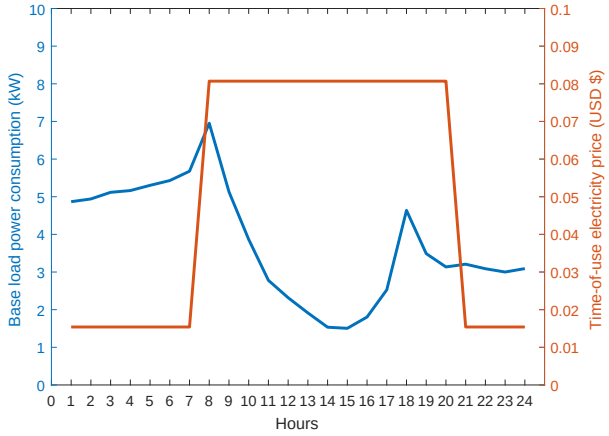

Figure 3: Base load power consumption and time-of-use electricity price

Table 2: Total operating cost for different control algorithms ($USD)

Method	Location 1	Location 2	Location 3	Average
Rule based	401.76	402.46	216.07	340.10
RLbEMS	389.36	392.4	208.95	330.24
NFQEMS	359.29	385.64	206.15	317.27
DQNEMS	358.36	386.26	205.22	316.61

Table 3: Daily average operating cost for different control algorithms ($USD)

Method	Location One		Location Two		Location Three	
	Mean	St.dev	Mean	St.dev	Mean	St.dev
Rule based	12.96	3.52	12.99	3.52	6.97	2.36
RLbEMS	12.56	3.41	12.66	3.42	6.74	2.29
NFQEMS	11.59	2.67	12.44	3.41	6.65	2.19
DQNEMS	11.56	2.67	12.46	3.42	6.62	2.29

Table 4: Daily average peak power consumption for different control algorithms (kW)

Method	Location One		Location Two		Location Three	
	Mean	St.dev	Mean	St.dev	Mean	St.dev
Rule based	16.42	1.90	17.00	2.55	13.08	1.72
RLbEMS	15.30	1.69	16.32	2.38	12.29	1.51
NFQEMS	11.02	1.90	11.63	2.55	8.78	1.74
DQNEMS	11.12	1.92	11.41	2.73	9.35	1.78

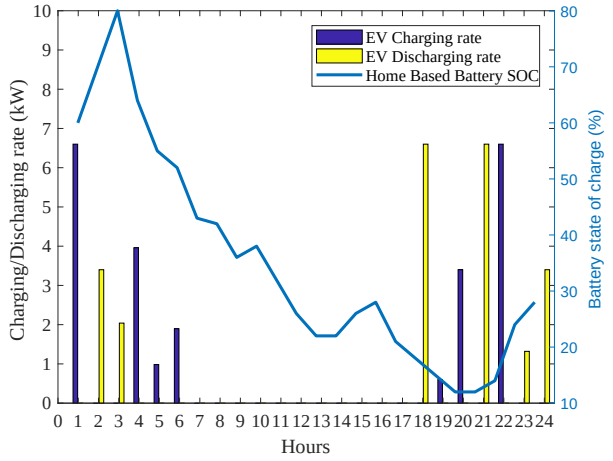

Figure 4: Home based battery and EV battery energy state (for location 1)

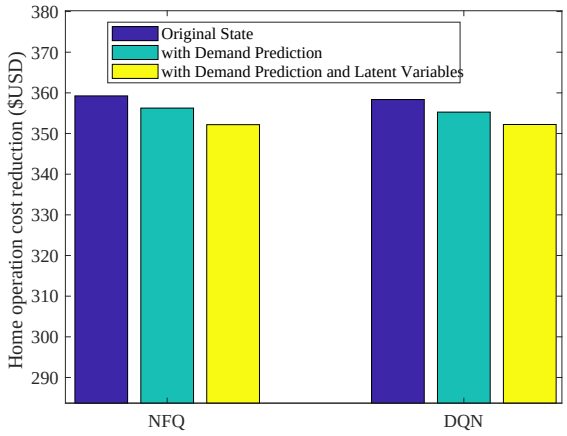

Figure 5: Total operating cost with RNN predictions

Peak power management is crucial for smart grid stability. Table 4 shows the daily peak power consumption for all three houses with different control algorithms. We can see that with the proposed control algorithms, daily peak power consumption can be significantly reduced. For NFQEMS it can reduce average daily peak power consumption by 32.39% over rule-based control and by 29.05% over RLbEMS control. For DQNEMS it can reduce by 31.42% of average daily peak power consumption over rule-based control and by 27.69% over RLbEMS control. In the smart home MDP formalization, peak power is not part of the reward function. Meanwhile, electricity price structure will influence the immediate reward. This means that NFQEMS and DQNEMS can successfully learn the TOU price structure. It also suggests that TOU price structure will encourage cost-sensitive customers to shift their deferable load to off-peak hours.

Figure 4 shows the EV charging, discharging and home battery energy state for house in location 1 under DQNEMS control. We can see that most EV charging is postponed from peak-hours to late in the night. After EV arrival, home EMS will choose to discharge EV battery and then charge the energy back when electricity price is lower. This can help reduce the peak power consumption and operating cost. However, the charging scheduling shown in Figure 4 still need to be improved. The discharging at 9pm and charging at 10pm may lead to potential power loss. An even better policy may be obtained with a finer discretization of the action space or with a continuous action space.

5.3 Enriching State Features with RNN Predictions

Figure 5 shows the operating cost for house in location 1 with original state features and with additional features of RNN predictions: load predictions and latent representations under control of NFQEMS and DQNEMS. It can be observed that when more related useful information are added to the input, we could further reduce the total operating cost.

6 CONCLUSION AND FUTURE WORK

With recent progress in advanced meters development and the prevalence of distributed energy generation, more attention has been paid to smart residential home energy management. Meanwhile, the increasing adoption of EVs brings new challenges and opportunities for smart grid energy management. In this paper, we show that the home energy management with EV charging can be formulated as an MDP. We propose two reinforcement learning based control algorithms: NFQEMS and DQNEMS to address it and show that home energy management can be dealt with both batch RL (off-line RL algorithm) and DQN (on-line RL algorithm). To use on-line RL algorithms, we need a simulator to interact with. We describe how to build a simulator with given historical data and battery specifications. Experiments based on real-world data show that the proposed two methods could significantly help reducing operating cost as well as peak power consumption over two baselines. We further use the predictions of RNN and show that both the demand prediction and latent representations of RNN could help improve the performance of the proposed control algorithms. In this paper, we use a feedforward neural network to approximate the Q value function and use discretized control actions. Our future work will investigate Deep RL algorithms with different neural network structures and continuous control actions for smart grid energy management problems.

The work in this paper, using discretized control actions, may encounter dimensionality issues with large and continuous action spaces. Reinforcement learning with continuous control actions could help relieve the dimensionality curse discussed. Besides, most of recent deep reinforcement learning successes are only in single agent domains while many real-world applications would involve interactions between different agents and require large-scale distributed control. Large-scale and complex decision-making problems are still very challenging for RL algorithms. Hierarchy Reinforcement Learning (HRL) could help to tackle these challenges. Smart grid is a complex electrical network. From a higher level, there are hierarchy structures for power systems and many power power systems are interconnected. At a lower level, for a certain neighborhood

level network, there maybe many residential houses in this network. If we only consider a single home, we may realize good peak power reduction for one specific home but we may not be able to guarantee that the whole power consumption would be reasonable. In the future, we plan to tackle large-scale smart grid energy management problems with HRL and continuous actions control.

REFERENCES

[1] Saima Aman, Yogesh Simmhan, and Viktor K Prasanna. 2013. Energy management systems: state of the art and emerging trends. *IEEE Communications Magazine* 51, 1 (2013), 114–119.

[2] Heider Berlink and Anna HR Costa. 2015. Batch Reinforcement Learning for Smart Home Energy Management.. In *IJCAI*. 2561–2567.

[3] Adriana Chiş, Jarmo Lundén, and Visa Koivunen. 2017. Reinforcement Learning-Based Plug-in Electric Vehicle Charging With Forecasted Price. *IEEE Transactions on Vehicular Technology* 66, 5 (2017), 3674–3684.

[4] conEdison. 2018. [Online]https://www.coned.com/en/save-money/energy-saving-programs/time-of-use. (2018).

[5] Damien Ernst, Pierre Geurts, and Louis Wehenkel. 2005. Tree-based batch mode reinforcement learning. *Journal of Machine Learning Research* 6, Apr (2005), 503–556.

[6] Vincent François-Lavet, Damien Ernst, and Raphael Fonteneau. 2017. On overfitting and asymptotic bias in batch reinforcement learning with partial observability. *arXiv preprint arXiv:1709.07796* (2017).

[7] Vincent François-Lavet, David Taralla, Damien Ernst, and Raphaël Fonteneau. 2016. Deep Reinforcement Learning Solutions for Energy Microgrids Management. In *European Workshop on Reinforcement Learning (EWRL 2016)*.

[8] Geoffrey J Gordon. 1995. Stable function approximation in dynamic programming. In *Machine Learning Proceedings 1995*. Elsevier, 261–268.

[9] Robert C Green II, Lingfeng Wang, and Mansoor Alam. 2011. The impact of plug-in hybrid electric vehicles on distribution networks: A review and outlook. *Renewable and sustainable energy reviews* 15, 1 (2011), 544–553.

[10] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.

[11] Honda. 2018. [Online]http://automobiles.honda.com/alternative-fuel-vehicles. (2018).

[12] Matthew E Kahn and Ryan K Vaughn. 2009. Green market geography: The spatial clustering of hybrid vehicles and LEED registered buildings. *The BE Journal of Economic Analysis & Policy* 9, 2 (2009).

[13] Anandalakshmi Thevampalayam Kaliappan, Swamidoss Sathiakumar, and Nandan Parameswaran. 2013. Flexible power consumption management using Q learning techniques in a smart home. In *Clean Energy and Technology (CEAT), 2013 IEEE Conference on*. IEEE, 342–347.

[14] Willett Kempton and Jasna Tomić. 2005. Vehicle-to-grid power implementation: From stabilizing the grid to supporting large-scale renewable energy. *Journal of power sources* 144, 1 (2005), 280–294.

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

[16] Long-Ji Lin. 1993. *Reinforcement learning for robots using neural networks*. Ph.D. Dissertation. Fujitsu Laboratories Ltd.

[17] Yudong Ma, Francesco Borrelli, Brandon Hencey, Brian Coffey, Sorin Bengea, and Philip Haves. 2012. Model predictive control for the operation of building cooling systems. *IEEE Transactions on control systems technology* 20, 3 (2012), 796–803.

[18] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. 2013. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602* (2013).

[19] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. 2015. Human-level control through deep reinforcement learning. *Nature* 518, 7540 (2015), 529.

[20] University of Queensland. 2018. [Online]http://www.uq.edu.au/solarenergy/pv-array/weather. (2018).

[21] OPENEI. 2018. [Online]http://en.openei.org/doe-opendata/dataset. (2018).

[22] Martin L Puterman. 1994. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons.

[23] Martin Riedmiller. 2005. Neural fitted Q iteration–first experiences with a data efficient neural reinforcement learning method. In *European Conference on Machine Learning*. Springer, 317–328.

[24] Stijn Vandael, Bert Claessens, Damien Ernst, Tom Holvoet, and Geert Deconinck. 2015. Reinforcement learning of heuristic EV fleet charging in a day-ahead electricity market. *IEEE Transactions on Smart Grid* 6, 4 (2015), 1795–1805.

[25] Qinglong Wang, Xue Liu, Jian Du, and Fanxin Kong. 2016. Smart charging for electric vehicles: A survey from the algorithmic perspective. *IEEE Communications Surveys & Tutorials* 18, 2 (2016), 1500–1517.

[26] Di Wu, Haibo Zeng, and Benoit Boulet. 2014. Neighborhood level network aware electric vehicle charging management with mixed control strategy. In *Electric Vehicle Conference (IEVC), 2014 IEEE International*. IEEE, 1–7.

[27] Di Wu, Haibo Zeng, Chao Lu, and Benoit Boulet. 2017. Two-Stage Energy Management for Office Buildings With Workplace EV Charging and Renewable Energy. *IEEE Transactions on Transportation Electrification* 3, 1 (2017), 225–237.

[28] Jian Xiong, Di Wu, Haibo Zeng, Shichao Liu, and Xiaoyu Wang. 2015. Impact assessment of electric vehicle charging on hydro Ottawa distribution networks at neighborhood levels. In *Electrical and Computer Engineering (CCECE), 2015 IEEE 28th Canadian Conference on*. IEEE, 1072–1077.