Balancing between Electric Vehicle Charging Station Income and Users Cost using Reinforcement Learning

Shaju Saha*, Abrar Zahin*, Nicholas Flann*
*Utah State University, Logan, UT
E-mail: ujas093017@gmail.com, abrarzahin303@gmail.com, nick.flann@gmail.com

***Abstract*—Increasing growing of electricity demand and environmental issues bring huge incentives to electric vehicles (EVs) market. EVs will improve the functionalities of present power system. On the other hand, unscheduled high penetration of EVs may have detrimental effects on power system performance. This project studies the electric EV charging scheduling problem under a charging station scenario, aiming to offer an optimal policy to optimize the battery configuration based on load prediction. Different from most existing works, we develop a charging scheduling based on Reinforcement Learning (RL) approach incorporating the practical battery charging characteristic, and design an intelligent charging management mechanism to maximize the interests of both the customers and the charging operator. These studies demonstrate that RL driven policy performs better in maximizing profit for EV charging station.**

Electric Vehicle, Reinforcement Learning, Markov Decision Process

I. Introduction

International commitments to reduce carbon dioxide (CO_2) emissions- the most common and pervasive greenhouse gashas fuelled efforts to decarbonize the traditional transport sector. Nowadays electric vehicles (EV) has become a reasonable and acceptable alternative to the conventional fossil fuelled vehicle. With the gradual increment of EVs the power load profile in distribution networks are prone to significant change. This in order to accommodate more clean energy, to reduce carbon emission and to alleviate peak charging loads convenient and publicly feasible EV charging infrastructure is needed.

Whereas, most existing literature has designed their charging scheduling model based on the assumption that future EV arrivals and electricity prices are known to charging stations when pricing and scheduling decisions are made.

In recent years, numerous day-ahead scheduling approaches have been proposed for this problem [1],[2]. For instance, in order to handle the uncertainty in electricity price, [1] developed a robust optimization approach for residential EV charging scheduling. Similarity, [3] proposed an information-gap-decision based approach to deal with the uncertainty in electricity price and optimize day-ahead scheduling of EV fleet. In [4], [5], EV fleet was formulated as a probabilistic virtual battery model, and scenario-based robust approaches were proposed to deal with the uncertainty of the EV users commuting behavior and the balancing requests. [6] studied the day-ahead scheduling of battery swapping stations where the uncertainty of the battery demand and the electricity price was modeled by inventory robust optimization and multi-band robust optimization, respectively Due to the existence of randomness in traffic conditions, users commuting behavior, and pricing process of the utility, EV arrival and departure time, EV energy consumption, and electricity prices are dynamic and time-varying. Therefore, efficiently managing EV charging/discharging to reduce the cost becomes challenging.

Real-time scheduling strategies that can respond to dynamic charging demand and time-varying electricity prices have attracted a lot of attention recently. For example, [7] developed a binary programming-based strategy to coordinate multiple EVs charging in a parking station in response to real-time curtailment request from the utility. [8] offered a formulation for the coordinated charging problem which considered the plug-in and plug-off frequency. Then, a real-time greedy algorithm is designed to.

Recently, model-free approaches which do not need any system model information have achieved great success in complex decision-making application [9]. This success has inspired the development of model-free approaches for smart grid applications [10],[11]. Compared to model-based approach, the advantage of the model-free approach is that it can learn a good control policy based on reinforcement learning (RL) and does not rely on any knowledge of the system [11].

Neural network has the potential of being universal approximator [12] and has been widely used for RL [13],[14]. In recent years, deep neural network achieved promising results in learning complex mapping from high-dimensional data. By utilizing deep neural network, deep RL has obtained significant success in many complex decision-making applications. For instance, a deep Q-network has achieved a level comparable to that of a professional human in the Atari 2600 [9]. LSTM has already showed promising result in predicting sequential information because of their capability to exploit long term dependencies among different sequence. However, to the best of our knowledge, application of LSTM in RL structure for an optimal policy in realtime EV charging/discharging problem has not been reported in the literature.

In this paper, the EV charging/discharging scheduling problem is formulated as value-iteration algorithm from the charging stations perspective. The objective is to find an optimal charging/discharging policy to take full advantage of the predicted real-time demand while fulfilling users driving demand. A model-free approach is proposed to determine the optimal schedules in a real-world scenario based on the deep RL. The proposed approach uses the predicted electricity prices in a specific time slot and battery State Of Charge (SOC) in that time slot as inputs, and outputs real-time charging/discharging polices. Unlike the traditional model-based methods, the proposed approach does not require any external system model information. Numerous experimental results demonstrate the effectiveness of the proposed approach.

The contributions of this paper are fourfold.

- Long Short Term Memory (LSTM) has been applied for predicting the demand with real-word data-set.
- A RL based approach has been proposed in order to design an optimal policy using the predicted demand to maximize the profit from charging station perspective.
- Based on the predicted demand an exponential pricing policy has been proposed in order to give customer flexibility to shift their demand in the lower price zone.
- Finally, based on the optimal scheduling policy profit/loss has been estimated for different battery size deployed in charging station for a period of one month.

II. Problem Definition

We formulate the real-time EV charging/discharging scheduling problem from the charging stations perspective. The time interval between two adjacent steps is one time slot, i.e. *on-peak, mid-peak* and *off-peak*. At time slot t, we observe the system state s_t which includes the information about the remaining charge in the charging station's battery and the past 48-hour electricity prices. Based on this information, we will choose the charging/discharging action a_t. This action represents the amount of energy that the station battery will charge or discharge during this time interval. After executing this action, we can observe the new system state s_{t+1} and choose the new charging/discharging action a_{t+1} for time step $t + 1$. Thus to summarize we can define our problem as

"*Given a fixed battery capacity in a charging station, predicted time-series load and real-time energy price, an optimal battery configuration for the EV station has been determined in order to decide whether to sell electricity to customer or to buy electricity from grid such that profit of charging station is maximized.* ". Finally, the proposed optimal policy has been enumerated for all possible battery sizes, which can be deployed in a charging station in order to observe the optimal profit for a specific charging station.

III. Proposed Approach

It is difficult to analytically determine the optimal policy π^* since the future electricity prices and users commuting behavior are unknown. A reinforcement learning (RL) solution is to iteratively update the value function $Q(s, a)$ based on demand prediction and proposed RL strucuture.

The LSTM network extracts discriminative features from the electricity price. After concatenating these features with battery SOC, the concatenated features are fed into a Q network to approximate the action-value of all feasible schedules under the given time slot. The schedule with the largest action-value is selected as the EV charging/discharging schedule. In the following subsections all necessary models, i.e LSTM, Value Iteration for RL, Training RNN, EV charging scheduling policy, Pricing Policy, Profit estimation of different battery size of our porpoised approach will be presented.

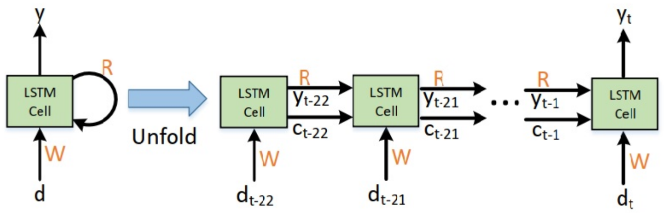

Fig. 1. An Unfolded LSTM Network

A. Long Short Term Memory (LSTM)

Extracting discriminative features from the raw data is a crucial step to improve the value function approximation. Good features should contain the information about the charging demand trends. With these features, the scheduling policy can maximize the profit for a charging station. In this paper, a LSTM network is proposed to extract these features, i.e charging demand trend.

Since electricity price fluctuates in a periodic way and has a natural temporal ordering, it is reasonable to infer future price trends from past electricity prices. Long Short-Term Memory (LSTM) network is known for its strong ability to model the time dependencies of time-series data [39], [40], and has achieved promising results in smart grid applications, such as load forecasting [16], [17].

The idea behind the LSTM network is to make use of sequential information, such as the real-time charging demand. LSTM network performs the same processing for every element of the sequence, with the output being dependent on the previous computations. The information about what has been The idea behind the LSTM network is to make use of sequential information, such as the real-time electricity prices. LSTM network performs the same processing for every element of the sequence, with the output being dependent on the previous computations. The information about what has been calculated so far can be stored or "*memorized*" in the LSTM cells. The typical structure of an LSTM network is shown in fig. (1).

The structure of the LSTM cell is shown in Fig. 3. The key to the LSTM network is the cell state c_t. The LSTM network has the ability to add information into or remove information from the cell state, carefully regulated by structures called gates. Gates are a way to optionally let information through. Specifically, an input gate determines the amount of information to be added into the cell state while a forget gate determines the amount of information to be inherited from the previous cell state c_{t1}. The input gate and forget gate are shown in Eq. (1) and (2) respectively,

$$i_t = \sigma(W_i * d_t + R_i * y_{t-1} + b_i) \tag{1}$$

$$f_t = \sigma(W_f * d_t + R_f * y_{t-1} + b_f) \tag{2}$$

where W_i, R_i, W_f , and R_f are the matrices of weights for the input gate and forget gate; b_i, and b_f are the vectors of biases for these gates; σ is the *sigmoid* function that outputs numbers between 0 and 1, describing how much of information should be let through.

The input information z_t is shown as

$$z_t = h(W_z * d_t + R_z * y_{t-1} + b_z) \tag{3}$$

where h denotes the *hyperbolic tangent* function; W_z and R_z are the matrices of weights; b_z is the vector of biases. The input gate would determine the amount of z_t to be added into the cell. Therefore, the cell state is calculated as

$$c_t = i_t \odot z_t + f_t \odot c_{t-1} \tag{4}$$

Where $\odot$ represents the element-wise multiplication operation; $i_t \odot z_t$ denotes the amount of information to be added from z_t; $f_t \odot c_{t-1}$ denotes the amount of information to be inherited from the previous cell state c_{t-1}. The output of the cell is determined by the output gate $o_t = \sigma(W_o * d_t + R_o * y_{t-1} + b_o)$, where W_o, R_o are the matrices of weights and b_o is the vector of biases. Thus, output of the cell can be shown by

$$y_t = o_t \odot h(c_t) \tag{5}$$

The output of the LSTM network, y_t, is concatenated with the battery SOC which is a scalar. These concatenated features, x_t, contain information about both the future price trends and the battery SOC. The information of the future price trends is essential to reduce the charging cost, while the information of the battery SOC is important to ensure the profit of the charging station will be maximized. Then, these concatenated features are fed into the Q network to approximate the optimal action-value function.

B. Value-Iteration for RL

Environment: Physical world in which the agent operates.
Action (A): All the possible moves that the agent can take.
State (S): Current situation returned by the environment.
Reward (R): An immediate return send back from the environment to evaluate the last action.
Policy(π): Policy is the strategy that the agent employs to determine next action based on the current state. Thus this policy function returns an action given a current environment state.

$$\pi(S) : S \rightarrow A \tag{6}$$

State transition model $(p(s_{t+1}|s_t, a_t))$: State transition model defines how the agent enters into a new state s_{t+1} from it's current state s_t having taken an action a_t. Reward Model $(p(r_{t+1}|s_t, a_t)$: Reward model describes the real number (*termed as reward*) that the agent receives from the environment after performing an action and entering to the next state. Discounting Factor (γ): It controls the importance of future rewards Value Function (V_s^π): The value function represents how good is a state for an agent to be in. It is expressed as expected total discounted reward agent get when starting from a state s and reach to the terminal state after vising all immediate states following a fixed policy π. Thus, for a given policy π to select actions, the corresponding value function is given by

$$V_s^\pi = E[\sum_{i=1}^{T} \gamma^{i-1} r_i | S_t = s] \quad \forall s \in S \tag{7}$$

Among all possible value-functions (*under different polices*), there exist an optimal value function (*optimal policy*) that has higher value than other functions for all states and thus, it is denoted by

$$V_s^* = max_\pi V_s^\pi \quad \forall s \in S \tag{8}$$

The optimal policy π^* is the policy that corresponds to optimal value function. So,

$$\pi^* = argmax_\pi V_s^\pi \quad \forall s \in S \tag{9}$$

C. Prediction of Charging Demand using LSTM

In order to predict the current time slot demand d_t, previous two day's *total demand, weekday or not, sine(day), cosine(day), demand at three previous time slot* d_{t-1}, d_{t-2} *and* d_{t-3} has been used as input features. Other parameters regrading the prediction model will be described in the result section.

D. Profit Maximization for EV stations

We designed our framework in order to maximize the profit for EV charging stations. For maximizing the profit we used value iteration mechanism which has been outlined in the previous section (III-B). The parameters related with our value iteration are provided below

1) State: Our state is consisted of { battery Charge Condition, hours in the day, load in that specific time }

a) Battery Charge Condition: Our battery condition in the EV stations are divided into 100 slots, each slot corresponding to remaining charge in percentage in the battery. For our problem we design our policy such that battery charge always remain in a certain limit in order to maximize the battery lifetime. We assume that battery charge of both $< 20\%$ and $> 80\%$ are injurious to the health of the battery.

b) Hours in the Day: As we are maximizing the savings at the end of the day, that's why we described our day in 24 discreet space, each space being considered as one hour.

c) Demand in a specific time-slot: From our data-set we saw that the demand increases according to the different peak hours. A visualization of such is given below

2) Action: Our action is consisted of { how much we are charging and how much we are discharging }

a) Charging Quantity: During taking action our RL agent can get charged whatever it needs.

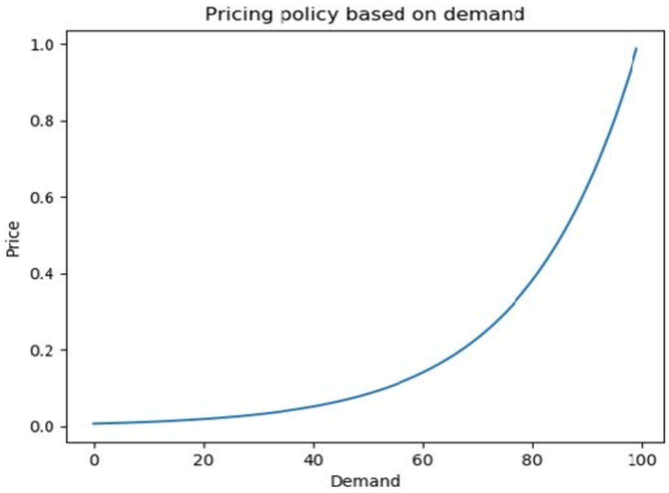

Fig. 2. Pricing Policy

b) Discharging Quantity: But during discharging RL agent discharges just according to the demand of the coming EVs.

3) Reward: We are considering reward as { how much EV stations are making at the end of the day}. For constructing the reward structure our corresponding assumptions are listed as below.

1) Penalize the agent if it charges the battery more than 80%.
2) Penalize the agent if it discharges more letting battery charges less than 20%.

E. Determining Pricing Policy

In our RL environment we do not sell any energy to the national grid rather we limit our buying and selling just in charging and discharging respectively. In order to maximize our profit we followed the policy of varying the price according to the demand.

$$Price = e^{(\frac{Demand-Low}{High-Low})} \times unit\ price \tag{10}$$

The motivation of such pricing policy (refer to fig. 2)is to give user full flexibility to shift their demand into lower price zone to minimize their cost also which will further result in shifting load to the off-peak hours.

F. Cost Optimization for different battery size

Finally, our proposed optimal scheduling policy has been enumerated all possible battery sizes to be deployed at the charging station and the profit/loss has been evaluated for each battery size over one month. Detail analysis of this policy will be given in the result section.

IV. Experimental Analysis

We evaluate the performance of our model with a confidential dataset. It contains 7 years data of numbers of charging station for EV in Utah, which and what type of vehicle are attending those charging stations, charging interval, charging port and charging quantity of each vehicle and energy consummations in each instance.

A. Data-set

The Dataset [18] includes energy consumption of each electric vehicle of different types, for each charger in every station.

B. Dataset Pre-processing

In order to better understand the overall scenario we divided each day into *on-peak, mid-peak* and *off-peak* hours. *On-peak, mid-peak* and *off-peak* are seen to be the following from our data-set for winter and summer seasons respectively.
For summer seasons.

1) *on-peak:* 8 am - 11 am, 6 pm - 7 pm
2) *mid-peak:* 12 pm -5 pm
3) *off-peak:* 12 am - 7 am 8 pm -12 pm

For winter seasons

1) *on-peak:* 12 pm - 5 pm
2) *mid-peak:* 8 am - 11 am, 6 pm -7 pm
3) *off-peak:* 12 am - 7 am, 8 pm -12 pm

For both seasons, whole day of Saturday and Sunday are considered as *off-peak.*

C. LSTM Prediction of current demand

For this part, we segmented every instances of each station so that we can use it as a station-wise training dataset to train our prediction model. For the prediction model we use previous one year dataset to train the model. Our dataset is designed to provide us demand on three kind of time slot, i.e. *on-peak, mid-peak* and *off-peak.* LSTM prediction model is designed to give the demand prediction of next time slot. For reader's convenience feature structure and output of the model are given below.

Feature = Previous one year's [*total demand in on-peak, mid-peak and off-peak , weekday or not, sine(day), cosine(day), demand at time slot* $t-1$, *demand at time slot* $t-2$].
Output = total demand in current time-slot

$$cosine(day\,of\,the\,year) = cos\frac{(2\times\pi\times i)}{365} \tag{11}$$

$$sine(day\,of\,the\,year) = sin\frac{(2\times\pi\times i)}{365} \tag{12}$$

Fig. (4) and fig. (5) is a visualization of demand as a function of day and time for a station in SLC and prediction for next time slot using demand from previous three time slots respectively. It can be concluded that even after having a large amount of randomness in the available real-world dataset (fig. 4) our prediction model performed quite well (fig. 5) reaching an accuracy of 68%.

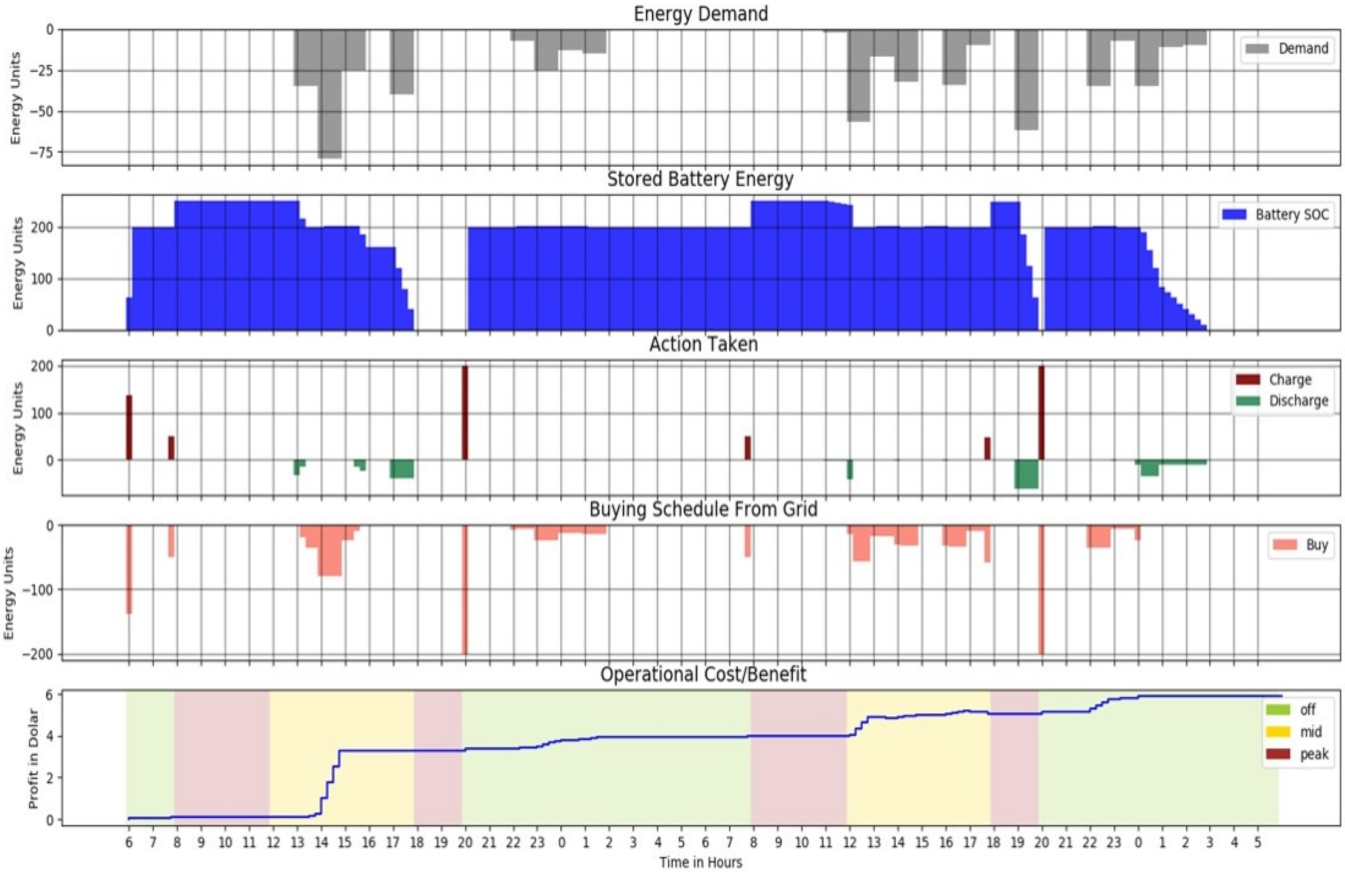

Fig. 3. Charging/ Discharging Scheduling for EV

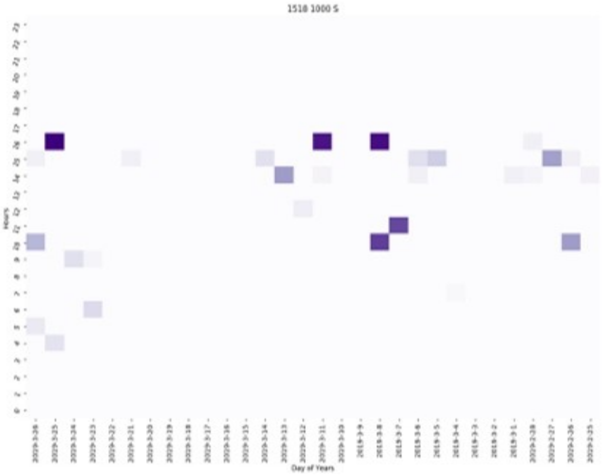

Fig. 4. Randomness of Data

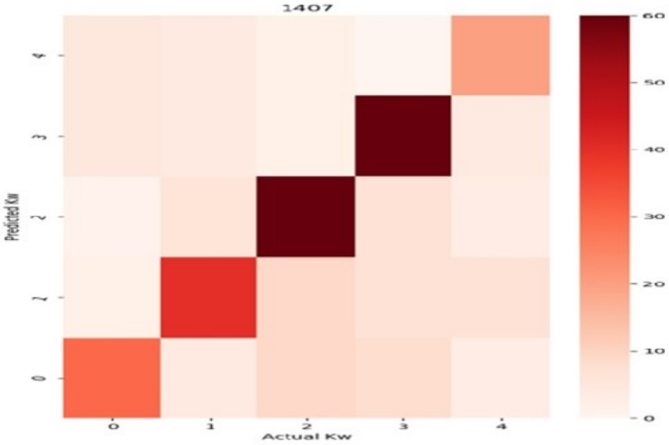

Fig. 5. Prediction heatmap

D. Optimal Charging/Discharging Schedule for EV using Value Iteration

Fig. (3) answers our following queries provided through optimal policy scheduling for EV charging station.

1) How much demand charging station expect at each time slot for two days?
2) What Battery state of charge (SOC) at each time slot should be maintained?
3) In order to maintain that SOC at each time slot what action (charging/discharging) need to be performed?
4) To perform that action how much electricity need to be

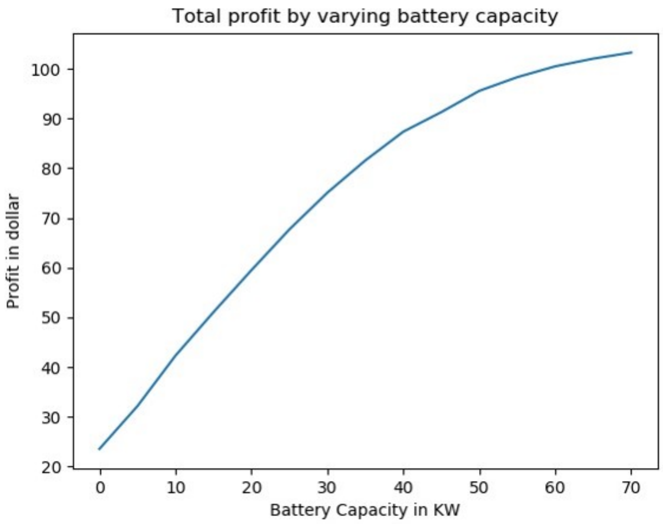

Fig. 6. Total Profit by Varying Battery Capacity

bought from grid?

5) What is the cumulative profit at each time slot?

respectively. A careful attention to the Fig. (3) gives us following observation

1) Battery tends to buy electricity exhaustively at off-peak hour in order to exploit the lower cost at off-peak time slot.
2) Battery tends to sell electricity from its storage during high demand because of our proposed pricing policy, to maximize the profit.

E. Cost Optimization for different battery size:

Finally, The proposed RL agent enumerated possible battery sizes to be deployed at the charging station and the profit/loss has been evaluated based on optimal scheduling, for each battery size over one month in fig. (6). The observation in (6) can be outlined as follows

1) Total profit increases with the size of the battery.
2) After a certain capacity total profit doesnt increase that much.
3) If battery size $\geq$ a days demand, agent tends to buy all the demand in off-peak hour and sell all day along. For this reason, after meeting that certain capacity profit tends to become more stable.

V. Conclusions

The primary motivation of this project was to develop an optimal policy based on RL in order to optimize cost from both customer and charging station perspective. In the near future, with the increasing availability of more dataset related with charging station this optimal policy will play a crucial role in balancing between profit maximization of charging station and users satisfaction.

References

[1]. M. A. Ortega-Vazquez, Optimal scheduling of electric vehicle charging and vehicle-to-grid services at household level including battery degradation and price uncertainty, IET Generation, Transmission Distribution, vol. 8, no. 6, pp. 10071016, June 2014.

[2] O. Erdinc, N. G. Paterakis, T. D. P. Mendes, A. G. Bakirtzis, and J. P. S. Catalo, Smart household operation considering bi-directional ev and ess utilization by real-time pricing-based dr, IEEE Transactions on Smart Grid, vol. 6, no. 3, pp. 12811291, May 2015.

[3] J. Zhao, C. Wan, Z. Xu, and J. Wang, Risk-based day-ahead scheduling of electric vehicle aggregator using information gap decision theory, IEEE Transactions on Smart Grid, vol. 8, no. 4, pp. 16091618, July 2017.

[4] M. G. Vaya and G. Andersson, Optimal bidding strategy of a plug-in electric vehicle aggregator in day-ahead electricity markets under uncertainty, IEEE Transactions on Power Systems, vol. 30, no. 5, pp. 23752385, Sept 2015.

[5] Self scheduling of plug-in electric vehicle aggregator to provide balancing services for wind power, IEEE Transactions on Sustainable Energy, vol. 7, no. 2, pp. 886899, April 2016.

[6] M. R. Sarker, H. Pandi, and M. A. Ortega-Vazquez, Optimal operation and services scheduling for an electric vehicle battery swapping station, IEEE Transactions on Power Systems, vol. 30, no. 2, pp. 901910, March 2015.

[7] L. Yao, W. H. Lim, and T. S. Tsai, A real-time charging scheme for demand response in electric vehicle parking station, IEEE Transactions on Smart Grid, vol. 8, no. 1, pp. 5262, Jan 2017.

[8] G. Binetti, A. Davoudi, D. Naso, B. Turchiano, and F. L. Lewis, Scalable real-time electric vehicles charging with discrete charging rates, IEEE Transactions on Smart Grid, vol. 6, no. 5, pp. 22112220, Sept 2015.

[9] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski et al., Human-level control through deep reinforcement learning, Nature, vol. 518, no. 7540, pp. 529533, 2015.

[10] Z. Wen, D. ONeill, and H. Maei, Optimal demand response using device-based reinforcement learning, IEEE Transactions on Smart Grid, vol. 6, no. 5, pp. 23122324, Sept 2015.

[11] F. Ruelens, B. J. Claessens, S. Vandael, B. D. Schutter, R. Babuska, and R. Belmans, Residential demand response of thermostatically controlled loads using batch reinforcement learning, IEEE Transactions on Smart Grid, vol. PP, no. 99, pp. 111, 2017.

[12] G. Cybenko, Approximation by superpositions of a sigmoidal function, Mathematics of Control, Signals, and Systems (MCSS), vol. 2, no. 4, pp. 303314, 1989.

[13] W. T. Miller, R. S. Sutton, and P. J. Werbos, A Menu of Designs for Reinforcement Learning Over Time. MIT Press, 1995, pp. 6795.

[14] H. He, Z. Ni, and J. Fu, A three-network architecture for on-line learning and optimization based on adaptive dynamic programming, Neurocomputing, vol. 78, no. 1, pp. 313, 2012.

[15] R. Bellman, Dynamic programming. Princeton Univeristy Press, John Wiley Sons, 1958.

[16] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, Short-term residential load forecasting based on lstm recurrent neural network, IEEE Transactions on Smart Grid, vol. PP, no. 99, pp. 11, 2017.

[17] W. Kong, Z. Y. Dong, D. J. Hill, F. Luo, and Y. Xu, Short-term residential load forecasting based on resident behaviour learning, IEEE Transactions on Power Systems, vol. PP, no. 99, pp. 11, 2017.

[18] http://www.na.chargepoint.com/