OPTIMIZATION OF PLUG-IN ELECTRIC VEHICLE CHARGING WITH FORECASTED PRICE

Adriana Chiş, Jarmo Lundén, and Visa Koivunen

Department of Signal Processing and Acoustics, Aalto University, Finland
Email: adriana.chis@aalto.fi, jarmo.lunden@aalto.fi, visa.koivunen@aalto.fi

ABSTRACT

This paper proposes a new method for scheduling the charging of plug-in electric vehicle's (PEV) battery. The method is employed in the demand side management of smart grids and has the goal of reducing the cost of charging over a long time horizon. The problem of scheduling the PEV battery charging is modeled as a Markov decision process with unknown transition probabilities. A *fitted Q-iteration* batch reinforcement learning algorithm with kernel-based approximation of the value iteration is proposed for learning the transition dynamics and solving the charging problem. The solution is obtained based on the knowledge of the true day-ahead electricity prices and predicted prices for the second day ahead. Simulation results using true pricing data demonstrate cost savings of 8%-40% for the consumer.

***Index Terms*—** Plug-in electric vehicles, smart charging, cost reduction, demand side management, reinforcement learning.

1. INTRODUCTION

The development of new technologies for electric vehicle batteries, the zero fuel consumption and low carbon emissions will make the plug-in electric vehicles (PEVs) a highly important component of the future transportation systems. The PEV charging, on the other hand, will cause an increase of electricity consumption for many households and hence also an increase of their electricity costs. A high penetration of PEVs will also affect the reliability of the power grid through even more unbalanced load profiles. One approach to overcome such problems could be the adoption of demand side management programs for reducing the energy cost and balancing the load in the power system. Smart charging programs can take advantage of the electricity price variation and the load flexibility in order to shift the electricity usage to times when the electricity prices and load in the power grid are low. But in many cases the consumers are slow and reluctant when it comes to changing their electricity usage habits. A good incentive to convince consumers of adopting smart charging programs would be considerable reduction of their electricity costs.

Most of the existing methods for optimization of PEV charging address the problems of balancing the load in the power grid [1] and reducing the electricity costs [2]. There are also methods that optimize the charging of PEVs in order to alleviate congestion in the power grid [3] or explore the possibility of PEV participation to the integration of renewable resources in the power system [4].

In this paper we propose a novel charging strategy of the PEV's battery that aims at reducing the cost of charging over a long time horizon. The method takes advantage of the electricity price fluctuations between two consecutive days in order to reduce the cost of charging for the consumer. The proposed strategy not only schedules the charging according to pricing incentives provided by dynamic pricing, but also according to constraints imposed by the driving patterns of the user. The price-based scheduling techniques for charging the PEVs present in the literature consider the electricity price to be known or given ahead for the entire scheduling period [1]. Other methods build their own pricing schemes [2, 4]. These methods perform the scheduling over shorter time periods that are typically of 24 hours or less and use hourly-based scheduling of consumption. The charging strategy proposed in this paper takes advantage of the large battery capacities existing these days that do not necessarily need daily charging and perform the scheduling of the PEV based on daily time steps. The method operates in a two-day window and uses known day-ahead prices for the current day and predicted prices for the following day. The method also chooses the amount of energy to be charged according to the constraints imposed by the known driving patterns of the user in order to diminish the possibility of the battery to deplete. The scheduling scheme considers only the hours of the day when the car is at home and the electricity price is at its minimum.

In this work the problem of scheduling the PEV battery charging is formulated as a Markov decision process (MDP). This work represents a significant improvement of the work presented in [5]. In [5], the PEV charging problem was solved using a Sarsa reinforcement learning algorithm with eligibility traces [6]. A significant drawback of this method is that the learned state-action values of the problem are stored in lookup tables. Consequently, both the state variables and actions need to be discretized. Hence, the algorithm requires a large amount of time to reach convergence. The method proposed in this paper overcomes these problems. We propose a *fitted Q-iteration* batch reinforcement learning algorithm with kernel based approximation of the value iteration for solving the PEV charging problem. This method speeds up the convergence significantly and works with continuous values of the MDP state variables. In order to make the method suitable for solving the PEV charging problem, we also present a new way of defining the rewards involved in the learning stage. In this work, at each step of the MDP, the rewards are determined according to the electricity price, current battery charge level and consumption. For demonstrating the performance gains of the proposed PEV charging approach we use true electricity pricing data. The simulation results show that the cost of electricity for charging the PEV can be reduced with 8% up to 40% when using the method proposed in this paper.

The paper is organized as follows. Section 2 presents the system model that formulates the charging problem as a MDP. The kernel-based *fitted Q-iteration* algorithm that solves the charging problem is presented in Section 3. The numerical results that show the performance of the proposed method are presented in Section 4. Section 5 gives the conclusions.

2. SYSTEM MODEL

This section describes how the PEV battery charging problem is formulated as a MDP with unknown transition probabilities. A MDP is a mathematical formulation for modeling decision making in uncertain situations [7]. A MDP scenario involves a sequence of steps and a set of states that describe the environment. For each state there is a set of possible actions. An agent takes actions according to a policy. A real valued reward function gives a reward for each action taken in each state. A state transition function defines the transition probability from one state to another.

In the case of the PEV battery charging problem the time steps of the MDP are measured in days. The time horizon of the scheduling is divided into two consecutive days denoted by d_1 and d_2. At the beginning of day d_1 the utility company gives the real market electricity prices per kWh for the next 24 hours. We denote these hourly price vector by $\mathbf{p}_1$. The electricity prices for the next day, d_2, are estimated using a price prediction algorithm and are denoted by $\mathbf{p}_2$. A consumption parameter c indicating an estimated amount of energy in kWh that the vehicle is expected to consume during day d_1 is assumed to be known. Two vector parameters, $\mathbf{h}_1$ and $\mathbf{h}_2$, indicate the hours of the day during which the car is expected to be at home during the days d_1 and d_2, respectively. These two parameters could be either selected by the user or learned from the driving patterns.

At each time step of the MDP, the state space of the battery charging problem is composed of three variables. Let ω denote a variable defining the day of the week corresponding to d_1 and b a continuous variable associated with the current charge level of the battery at the beginning of d_1. The battery capacity is limited by a maximum value denoted by b_{max}. The third state variable Δ, which is continuous as well, describes the difference between the minimum hourly charging cost, if the charging would occur on day d_1, and the estimated minimum hourly charging cost for day d_2. This variable is calculated as:

$$\Delta_1 = \min_{h \in \mathbf{h}_1} p_{1,h} \cdot c_r - \min_{h \in \mathbf{h}_2} p_{2,h} \cdot c_r, \tag{1}$$

where by $p_{i,h}$ we denote the hth element of vector $\mathbf{p}_i$, *i=1,2*. Here c_r represents the hourly charging rate of the battery. The state of our problem at the beginning of a day d_1 is expressed as: $\mathbf{x} = [\omega \ b \ \Delta]$.

The action space of the PEV battery charging problem is discrete. The battery capacity is assumed to be discretized into *L* levels. An action is denoted by u and is defined by the operation that controls the charging of the battery. The control operation chooses the number of battery levels to be charged during day d_1, where $u \in \{0, ..., L-1\}$. We denote by $u_{\mathfrak{L}}$ the amount of energy in kWh corresponding to charging *u* battery levels.

The method that we propose for solving the PEV charging problem is a batch reinforcement learning method called *fitted Q-iteration* with kernel-based approximation of the value iteration [8].

3. FITTED Q-ITERATION WITH KERNEL BASED APPROXIMATION OF THE VALUE ITERATION

Fitted Q-iteration is a batch reinforcement learning method in which the unknown transition dynamics of the MDP are learned from an available batch of transition samples:

$$\mathfrak{F} = \{(\mathbf{x}_l, u_l, \mathbf{y}_l, r_l) | l = 1, ..., \#\mathfrak{F}\}. \tag{2}$$

Here $\mathbf{x}_l$ represents the current state , u_l is the action taken in the current state, r_l is the reward obtained for taking the action u_l in the state $\mathbf{x}_l$ and $\mathbf{y}_l$ the next state that the system reaches. By $\#\mathfrak{F}$ we denote the number of available samples in the set $\mathfrak{F}$.

The goal of batch reinforcement learning is to find a policy that maximizes the rewards received by an agent when taking actions in the current environment. Let π_N be a control policy corresponding to a temporal horizon of *N* steps. For a given initial state, the objective of our problem would be the maximization of the expected discounted return over *N* steps when the system is controlled by policy π_N:

$$J_N^{\pi_N}(\mathbf{x}) = \mathbb{E}\left\{\sum_{n=0}^{N-1} \gamma^n r_n | \mathbf{x}_0 = \mathbf{x}\right\}, \tag{3}$$

where γ represents the discount factor.

The purpose of the proposed algorithm is to approximate the *action-value* function of a policy π. This function is also called *Q-function* and represents the discounted return when taking an action in a specific state according to the policy π. At the end of the learning process, at each step *n=0,...,N-1*, the optimal control policy that will minimize the cost of charging is the one that satisfies:

$$\mu^* = arg \max_u Q(\mathbf{x}_n, u). \tag{4}$$

The kernel-based approximation of the value iteration was introduced in [9]. In order to obtain the kernel-based approximation of the Q action-value consider that $S_u = \{(\mathbf{x}_l, \mathbf{y}_l) | l = 1, ..., m_u\}$ is the collection of all historical state transitions from the set of samples $\mathfrak{F}$ where action *u* was taken. The kernel-based Q action-value update during the learning process is [9]:

$$\hat{Q}_{\tau+1}(\mathbf{y}'_l, u) = \sum_{(\mathbf{x}_l, \mathbf{y}_l) \in S_u} \kappa(\mathbf{x}_l, \mathbf{y}'_l)[r_l + \gamma \max_u \hat{Q}_\tau(\mathbf{y}'_l, u)], \tag{5}$$

where $\kappa(\mathbf{x}_l, \mathbf{x})$ is a weighting kernel function that depends on the training set $\mathfrak{F}$, the state $\mathbf{x}$ and a kernel function ϕ^+. The weighting kernel function has the following expression:

$$\kappa(\mathbf{x}_l, \mathbf{x}) = \frac{\phi^+\left(\frac{\|\mathbf{x}_l - \mathbf{x}\|}{\beta}\right)}{\sum_{\mathbf{x}_k \in \mathfrak{F}} \phi^+\left(\frac{\|\mathbf{x}_k - \mathbf{x}\|}{\beta}\right)}, \tag{6}$$

where β is a *bandwidth* parameter and controls the smoothness of the kernel function.

The learning stage is completed when the difference between the values of $\hat{Q}(\mathbf{y}'_l, u)$ from two consecutive iterations becomes almost zero. The value of $\hat{Q}(\mathbf{x}, u)$ for any new state $\mathbf{x}$ can be obtained from:

$$\hat{Q}(\mathbf{x}, u) = \sum_{(\mathbf{x}_l, \mathbf{y}_l) \in S_u} \kappa(\mathbf{x}_l, \mathbf{x})[r_l + \gamma \hat{Q}(\mathbf{y}_l, u)]. \tag{7}$$

We design the reward values *r*($\mathbf{x}$,*u*), needed for the learning stage, to indicate a desired choice of actions at each step of our MDP, when the scheduling of the PEV charging is performed. The purpose is to minimize the cost of PEV charging to the consumer and to prevent the battery from depleting. The increase or decrease of electricity prices between two consecutive days is indicated by the sign of variable Δ. A positive sign of Δ indicates a decrease in electricity prices and a negative sign of Δ shows an increase of electricity prices. The reward values *r*($\mathbf{x}$,*u*) are chosen according to the value of the pricing parameter Δ, the value of the battery state parameter b and the value of the consumption parameter c. Table 1 gives a description of how these values are chosen. The actual value of the reward in case of that particular state $\mathbf{x}$ and action u is then:

Table 1: REWARDS

$\Delta \geq 0$	
$b \leq c$	$r(\mathbf{x}, u) > 0$ if $u_{\mathfrak{L}} = u_{min}$
$b \geq c$	$r(\mathbf{x}, u > 0$ if $u_{\mathfrak{L}} = u_{min} = 0$

$\Delta < 0$	
$b \leq c$	$r(\mathbf{x}, u) > 0$ if $u_{min} + 1 \leq u_{\mathfrak{L}} \leq u_{max}$
$b \geq c$	$r(\mathbf{x}, u) > 0$ if $1 \leq u_{\mathfrak{L}} \leq u_{max}$

$$r(\mathbf{x},u) = \begin{cases} a & \text{if } u_{\mathfrak{L}} = u_{min} = 0, \\ z & \text{if } r(\mathbf{x},u) > 0 \text{ as in Table 1}, \\ 0 & \text{otherwise}, \end{cases} \tag{8}$$

where a and z are some real values proportional to the compensation that the agent is believed to deserve for taking the action action u in the state $\mathbf{x}$. Here u_{min} denotes the amount of energy in kWh corresponding to the minimum number of battery levels that must be charged during a day such that the battery does not deplete ($u_{min} \geq c$) and u_{max} denotes the amount of energy corresponding to the maximum number of battery levels that can be charged so that b+$u_{max} \leq b_{max}$.

4. PERFORMANCE EVALUATION

In order to evaluate the performance of the proposed PEV charging method we consider the case of a PEV with battery capacity b_{max}=24kWh. The range of the PEV is considered to be 160 km and the battery charging rate is c_r=6.6 kW. The variable ω, defining the day of the week when the scheduling occurs, takes a value $\omega \in \{1, ..., 7\}$. The daily battery consumption is simulated considering that during each day of the week the car is driven for a random distance with a mean between 10 and 60 km. The following mean distance values are considered in our simulations for each day of the week: $\{50, 50, 60, 55, 50, 25, 10\}$. A uniform random variation between 0 and $\pm$ 20 % is added to the calculated consumption. The pricing parameter Δ is limited to a continuous value within the interval [-15¢,15¢]. The calculated values for Δ which are less or greater than the limiting values of this interval are set to -15¢ and 15¢, respectively. The battery capacity is discretized into L=8 levels. The action chosen by the proposed scheduling method at the beginning of each day d_1 represents the number of battery levels that should be charged during that day: $u \in \{0,1,...,7\}$. The values a and z considered in the simulations for calculating the rewards $r(\mathbf{x},u)$ in (8) are: a=600 and z=$u \cdot 100$.

The simulations are performed using true pricing data taken from the ISO New England database [10]. We employ a Bayesian neural network [11] for obtaining the predicted prices. A set of hourly true and predicted prices prices for 630 days was selected starting from the 8th of August 2011 until the 28th of April 2013. This set was divided into 4 different subsets of hourly pricing data. Two subsets containing a total of hourly prices for 406 days were used for the learning stage and the remaining hourly prices for 224 days were used for testing. In order to increase the robustness of learning, the set of four-tuple samples $\mathfrak{F}$ needed for the learning stage was built of a mixed set of data composed by samples generated using true data prices and samples generated using simulated data prices. The set of samples containing the true prices was generated using several thousand episodes based on the considered set of prices. For building the simulated prices samples set we considered a collection of possible prices such that the pricing parameter Δ would be within the interval [-15¢,15¢]. To these prices we assigned a set of possible battery state variables within the interval [0, b_{max}]. The actions taken at each step were randomly chosen among the set of actions that will always keep the battery charge level within [0, b_{max}]. The *Q action-value* function is computed separately for each day of the week. In order to build the subsets S_u a number of m_u=3500 samples were chosen randomly for each action. The data was then normalized such that the state variables take values between 0 and 1.

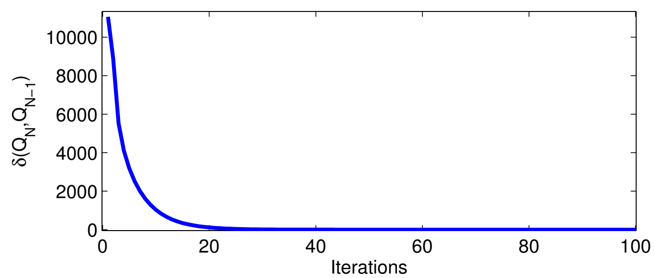

Fig. 1: The mean square distance between the approximated $\hat{Q}$ values from consecutive iterations shows that the proposed algorithm reaches convergence after about 30 iterations.

The parameters of the *fitted Q-iteration* algorithm used for solving of the PEV charging problem are presented in Table 2. These parameters were experimentally found most suitable among a set of different kernel functions, distance norms and different values for the bandwidth parameter. The weighting kernel function $\kappa(\mathbf{x}_l, \mathbf{x})$ in (6) is computed as the product kernel of the function corresponding to the battery state variable b and the kernel function corresponding to the pricing parameter $\Delta : \kappa(\mathbf{x}_l, \mathbf{x}) = \kappa(b_l, b) \cdot \kappa(\Delta_l, \Delta)$.

Table 2: SIMULATION PARAMETERS

Discount factor γ	0.9
Kernel function $\phi^+(\cdot)$	$e^{-(\cdot)}$ -Laplacian kernel
Bandwidth parameter β_{κ_b}	0.01
Bandwidth parameter β_{κ_Δ}	0.01
Norm distance $\|\| \cdot \|\|$	$\| \cdot \|$ -Absolute distance

The the mean squared distance between the approximated $\hat{Q}$ values from consecutive iterations was employed for evaluating the convergence of the proposed algorithm: $\delta(\hat{Q}_N, \hat{Q}_{N-1}) = \sum_{l=1}^{\#\mathfrak{F}} (\hat{Q}_N(\mathbf{x}_l, u_l) - \hat{Q}_{N-1}(\mathbf{x}_l, u_l))^2 / \#\mathfrak{F}$. The variations of this value over 100 iterations are presented in Figure 1. It is noticeable that this criterion reaches a value close to zero after roughly 30 iterations. This means that on average the approximated Q state-action values reach convergence after roughly 30 iterations. For obtaining an accurate convergence of all state-action values of the Q function it is recommended to run the algorithm a little longer.

The evaluation of the proposed charging policy is performed using the policy obtained after running the *fitted Q-iteration* algorithm for 100 iterations. The charging takes place during the hours of the day d_1 when the car is at home, given by the vector $\mathbf{h}_1$. The total cost of charging the PEV battery during the day d_1 is calculated as:

$$p(d_1) = \sum_{j=1}^{\epsilon} g_j \cdot c_r + g_{\epsilon+1} \cdot c_r \cdot c_\epsilon, \tag{9}$$

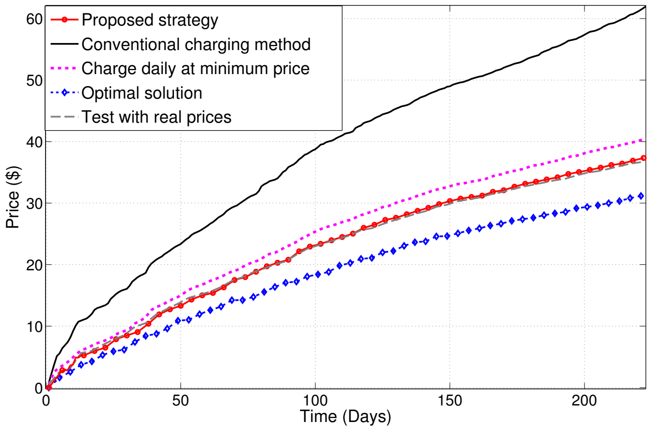

Fig. 2: Using the proposed strategy 40% cost reduction is obtained in comparison with the conventional strategy and a 8% reduction is obtained in comparison with the smart charging strategy. The test scenario with real prices only indicates that our solution is robust to the price prediction errors. The charging costs need to be further reduced with 15% in order to reach the global optimum solution.

where $\epsilon = \left\lfloor \frac{u_{\mathfrak{L}}}{c_r} \right\rfloor$, $c_\epsilon = \frac{u_{\mathfrak{L}}}{c_r} - \left\lfloor \frac{u_{\mathfrak{L}}}{c_r} \right\rfloor$ and $\mathbf{g} = [g_1, \ldots, g_{\#\mathbf{h}_1}]^T$ is the vector of hourly prices ordered in ascending order in which $g_i = p_{1,h_{1,i}}, i = 1, \ldots, \#\mathbf{h}_1$. By $\#\mathbf{h}_1$ we denote the length of the vector $\mathbf{h}_1$ and $h_{1,i}$ is the value of its ith element.

Figure 2 shows the cumulative charging cost over 224 days when the proposed PEV charging scheduling method in employed in comparison with other PEV charging policies. The performance curve of the proposed strategy represents the average value over 20 different car usage realizations. These realizations were performed using different car consumption values generated according to the randomized process explained earlier. The conventional charging policy fully charges the battery every day at random hours without considering the fluctuations of electricity price. A smart charging strategy, considered for comparison purpose, implies that the driver fully charges the battery every day during the hours when the price is minimum. Another plotted curve shows a lower bound defined by the cumulative price of charging when the global optimal charging strategy would be used. This strategy considers the hypothetical situation when the true electricity price and the car consumption would be known for the entire time span of 224 days. The figure shows as well a scenario for testing the proposed strategy when true electricity prices are used also for the second day ahead, instead of the predicted prices. We can observe from the plots that the gaps between the charging costs increase as a function of time. Roughly a 40% decrease in charging costs could be observed when the proposed method is used instead of the conventional charging policy. Even if the battery is fully charged daily at minimum price of the day, the driver still saves money when using the novel strategy. The reduction in the cost of charging being about 8% in comparison with the other smart charging strategy. For reaching the global optimal solution the charging cost of the proposed strategy would still have to be decreased by another 15%, but we can still say that our strategy that uses one day predicted prices brings us significantly closer to the global optimal solution. Observing the solution of the test scenario with true prices only, we can see that this curve nearly overlaps the solution of the proposed strategy. This shows that the solution of the proposed charging strategy is robust to some degree of error in the price prediction.

5. CONCLUSION

A novel method for scheduling the EV battery charging has been proposed in this paper. The method operates in a two-day window and involves the knowledge of true day-ahead prices and predicted prices for the second day ahead. The goal of the proposed method for scheduling the PEV charging is to decrease the cost of charging over an infinite time horizon. The user's driving patterns are taken into account in the scheduling technique in order to diminish the probability of the battery to deplete. The problem of scheduling the PEV charging has been formulated as a MDP. In comparison to the previous work [5], a new method for solving the PEV charging problem has been proposed here. This method employs a *fitted Q-iteration* batch reinforcement learning algorithm with kernel based approximation of the value iteration. The main advantages of this new method are that it uses continuous values of the MDP state variables and converges fast to the solution. We also presented a new method of defining the rewards involved in the learning stage. The evaluation of the proposed PEV charging scheme was performed using true pricing data. Simulations show that the method reduces the charging price with a value between 8% and 40%.

6. REFERENCES

[1] N. Rotering and M. Ilic, "Optimal charge control of plug-in hybrid electric vehicles in deregulated electricity markets," *IEEE Trans. on Power Systems*, vol. 26, no. 3, pp. 1021–1029, August 2011.

[2] Y. He, B. Vankatesh, and L. Guan, "Optimal scheduling for charging and discharging of electric vehicles," *IEEE Trans. on Smart Grid*, vol. 3, no. 3, pp. 1095–1105, September 2012.

[3] R. Li, Q. Wu, and S. S. Oren, "Distribution locational marginal pricing for optimal electric vehicle charging management," *IEEE Trans. on Power Systems*, vol. 29, no. 1, pp. 203–211, January 2014.

[4] Z. Tan, P. Yang, and A. Nehorai, "An optimal and distributed demand response strategy with electric vehicles in the smart grid," *IEEE Trans. on Smart Grid*, vol. 5, no. 2, pp. 861–869, March 2014.

[5] A. Chiş, J. Lundén, and V. Koivunen, "Scheduling of plug-in electric vehicle battery charging with price prediction," in *Proc.4th IEEE Power and Energy Society Conference on Innovative Smart Grid Technologies Europe*, October 2013.

[6] R. Sutton and A. Barto, *Reinforcement learning*, MIT Press, 1998.

[7] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, John Wiley & Sons, Inc., 1994, 2005.

[8] L. Buşoniu, R. Babuška, B. De Schutter, and D. Ernst, *Reinforcement Learning and Dynamic Programming Using Function Approximators*, MIT Press, 2010.

[9] D. Ormoneit and Ś. Sen, "Kernel-based reinforcement learning," *Machine learning*, vol. 49, no. 2-3, pp. 161–178, 2002.

[10] www.iso ne.com.

[11] R. M. Neal, *Bayesian Learning for Neural Networks*, vol. 118, Springer-Verlag, 1996.