Study on Aggregator's Strategy of Controlling Electric Vehicles to Compensate Imbalances in Power Systems Using Reinforcement Learning

Ryuhei Furuta and Ryuji Matsuhashi
The University of Tokyo, Electrical Engineering Department, Tokyo, Japan
Email: {furuta, matu}@enesys.t.u-tokyo.ac.jp

***Abstract*— In Japan, electricity market became open to any companies in 2016. Due to this legal change, a lot of new power producer and supplier (PPS) could join this market and they aim to maximize their profits. Now, they have to consider the electricity control method because the imbalance penalty would increase in the future. In this paper, we suppose a control strategy of these aggregators considering this imbalance penalty. In this system, these aggregators can control each household's electric vehicle as a control device by paying some money to the EV owners. Of course, offering a higher reward to EV owners, they can use more electric vehicles to avoid the imbalance penalty than when they offer a lower reward. However, the too high reward is a heavy burden for these companies. Using reinforcement learning, the aggregator can learn the strategy of controlling electric vehicles without defining the households' precise information. As a result, we could achieve a 30% cost reduction compared to when no electric vehicles are used for reducing imbalance penalty.**

***Index Terms*—reinforcement learning, EV charging, imbalances, power systems**

I. Introduction

A. Background

In 2016, the electricity market became open to anyone in Japan. Nowadays, there are lots of emerging companies producing and selling electricity in Japan. Most of these new power producer and supplier (PPS) tend to only have renewables as a power generator and not a controllable generator such as a thermal power plant. Since most of them don't have a controllable generator, if they can't supply electricity precisely, they only rely on a general electricity utility by paying the imbalance penalty. This imbalance penalty is not so expensive now, however, it would get much higher in the future and be a heavy burden for these new PPS.

Recently, energy storage systems have attracted attention. In Japan, a lot of renewable energy is introduced to the power system, which may make the energy system relatively unstable. Energy storage systems, for example, battery, electric vehicle (EV) and demand response, are regarded as one of the solutions to the problem. New PPS are required to operate more responsibly using these technologies.

Reinforcement learning is also one of the noteworthy technologies. By using this technology, the aggregator can learn the best action based on the historical actions and results. The main feature of reinforcement learning is that it doesn't need a well-defined model because it can learn the best policy from its own action and result history.

B. Related Work

As pointed out in the last subsection, electric vehicles and reinforcement learning are one of the most attention-getting technologies. Thus, there are lots of studies regarding these two technologies. In this section, we explain some recent studies about these technologies.

Electric vehicles are used as a means of transportation, of course. However, at the same time, electric vehicles can be used as a battery when they are connected to the grid. Stijn *et al.* [1] paid attention to the usefulness of electric vehicles and considered the management method of EV fleets. The paper presented a learning schema of how much electricity the aggregator should buy in the day-ahead market. Jasna and Willett [2] evaluate how EV fleets are used for grid support by some case studies and simulated how much EVs can earn if they join the regulation market in the United States. They concluded that using EVs are cost effective and would also improve the stability of the electrical grid. Takeda *et al.* [3] actually did the experiment that tested the controllability of EV and confirmed that EV can get the control signal from PC. They also made clear that introducing EVs can stabilize grid frequency and effective for load frequency control. Junjie *et al.* [4] summarized the studies related to EV fleet management and classified EV fleets' services and control methods.

As for reinforcement learning, because of its wide range of applications, there are many studies using this technique to this academic field, that is, energy systems and economics. Daniel *et al.* [5] proposed a novel energy management system (Demand Response system) using reinforcement learning. Since people act differently based on their lifestyles, it is impossible to do optimization at any time at each home. Reinforcement learning can

Manuscript received May 13, 2019; revised November 15, 2019.

doi: 10.18178/ijoee.7.2.67-71

automatically adjust their lifestyles and reduce end-user financial costs from 16% to 40%. However, only one household's demand fluctuates drastically. Thus, Ivana *et al.* [6] proposed multi-agent demand response, which means demand response conducted by some households. They used reinforcement learning based on their load forecasting. Tiago *et al.* [7] proposed multi-agent reinforcement learning in electricity markets. With reinforcement learning, each agent can learn the best possible bids under the market situation. A.L. Dimeas and N.D. Hatziargyriou [8] applied these multi-agent learning to microgrids operations. Using reinforcement learning, each agent can learn and solve a decision-making problem without a central controller.

Regarding EV charging, Wenbo and Vincent [9] used reinforcement learning to find a better EV charging strategy under price uncertainty. In their paper, electricity cost can vary like the Markov chain, and with many times iterations, EV can learn when they should charge and sell electricity. Frederik *et al.* [10] added the uncertainty of arrival and departure times of each EV and proposed approximate dynamic programming. Their study showed their proposal method can reduce the aggregator's expensive peak charging and penalties that resulted from not supplying enough electricity to the consumer. Konstantina *et al.* [11] applied the individual utility function to EV charging. This is because how much risks they take depends on their way of thinking. For example, the risk-averse people don't want to keep their EVs not fully charged.

C. *The Purpose of the Study*

Considering these backgrounds, the achievement of this paper is proposing the aggregator's control strategy of EVs using reinforcement learning considering the trade-off relationship between reward to consumers and imbalance penalty to general electricity utility.

II. Simulation Model

A. *Aggregator Model*

The overall schema of this paper is shown in Fig. 1. The aggregator can control EV fleets, and, with EVs, the aggregator aims to minimize its imbalance penalty cost and paying rewards to each EV. Please note that EVs aren't owned by the aggregator in this system, thus the aggregator has to pay some revenue to each EV and the EV owners decide whether they supply electricity or not, considering the offered price. Regarding the offered price, we assume p = 0.5, 1, 2, 4 [yen/kWh], and the imbalance penalty is 24 [yen /kWh]. This is about three times as much as today's price. However, we assume the price gets much higher than the present one.

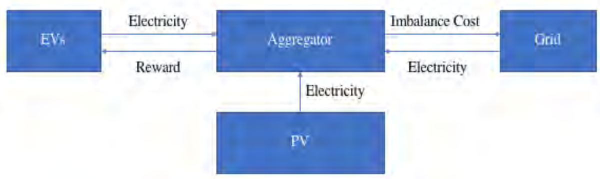

Figure 1. The schema of the aggregator's model.

The aggregator couldn't assume PV output precisely. If the aggregator predicts too wrong output, it would make imbalance penalty. In that sense, prediction of PV output is one theme of this field. [12] In this paper, we assume PV output fluctuates like Fig. 2 below. The red line means PV output prediction, and it may cause errors with a certain width like a blue line. This prediction is based on [13].

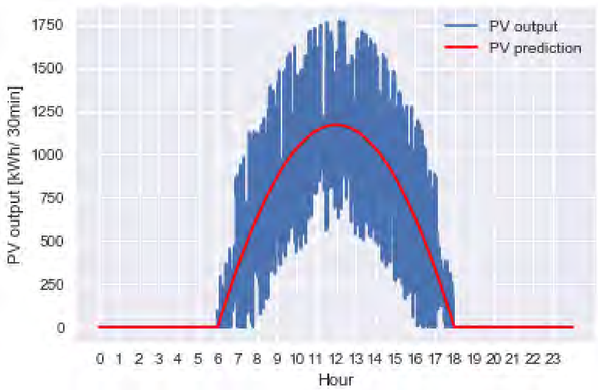

Figure 2. PV output and prediction error.

B. *EV Owner's Model*

EVs owned by households are not always available for electricity storage. In this paper, we assume that the availability of EVs is like Fig. 3. Please note that each time slot is fixed to 30 minutes.

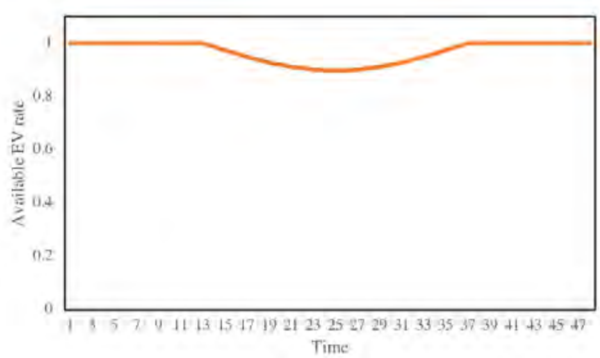

Figure 3. Available EV rate at each time slot.

Regarding EV owners' decision-making, we use logit model, which is famous for decision-making process. The equations are shown in (1), (2):

$$P1 = \frac{1}{1 + \exp(\eta(V2 - V1))} \tag{1}$$

$$P2 = 1 - P1 \tag{2}$$

V1 and V2 represent the consumer's utility when the decision maker chooses 1 (joining this aggregator's control at this time slot) or 2 (not joining this schema at this time slot), respectively. In the logit model, people make the V1 decision if V2 - V1 is low, that is, V1 is much more profitable than V2. The typical graph of this model is shown in Fig. 4.d

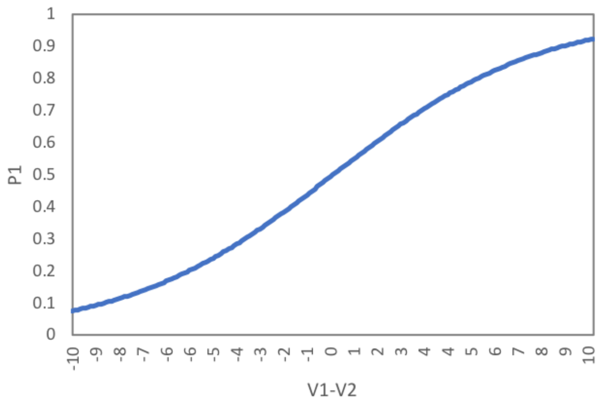

Figure 4. A typical graph of the logit model.

With the logit model, we simulated the decision-making process of EV owners, which means whether they should supply EV capacity or not at the offered price. Regarding η, we assume that η will become higher if the offered price gets higher than the ahead price, and vice versa. By introducing this idea, we could simulate reinforcement learning.

C. Reinforcement Learning

We use reinforcement learning to learn good EVs' reward price strategy by the aggregator itself. Q-learning is one of the most famous reinforcement learning algorithms. In this subsection, we explain the concept of reinforcement learning.

At first, reinforcement learning consists of an agent and environment. At each time, the agent takes the action that the agent thinks it is the best choice for the environment. Then, environment changes and the agent get some reward from the environment. Repeatedly doing this cycle, the agent can learn what it is the best action in the present environment. The signal that evaluates the result of the action is Q-value. Q-value is updated following the equation below:

$$Q(\mathbf{s}_t, a_t) = Q(\mathbf{s}_t, a_t) + \alpha\{r(\mathbf{s}_t, a_t) + \max Q(\mathbf{s}_{t+1}, a_t) - Q(\mathbf{s}_t, a_t)\} \tag{3}$$

α is a constant and it means learning rate, generally, it is set to a small number. In this paper, we set α to 0.001. **s** means the state vector that corresponds to each state and a means the action that the agent can take. In this paper, we assume that **s** consists of the following variables shown in Table I. The reward **r** means how much it would cost if the price is set to at value, that is the sum of the imbalance penalty cost and the reward to each EV owner.

TABLE I. THE MAIN VARIABLES COMPOSING STATE

Imbalance Compensation [kWh / 30min]	0, 100, 200, 300, 400, 500, 600, 700
Available EV rate (Shown in Fig.3)	0.9, 1.0
η	0.8, 1.4, 2.0

TABLE II. THE ACTION VALUES

Reward to EV [yen /kWh]	0.5, 1.0, 2.0, 4.0

The action consists of how much the aggregator pays to each EV. This variable is shown in Table II.

III. RESULT

First, to overview the general tendency, we set the number of EVs to 200 and did a simulation. The result is shown in Fig. 5. In Fig. 5, cost includes both imbalance cost and the reward to each household's EV.

$$\text{Total Cost} = \text{Imbalance Cost} + \text{EV Reward} \tag{4}$$

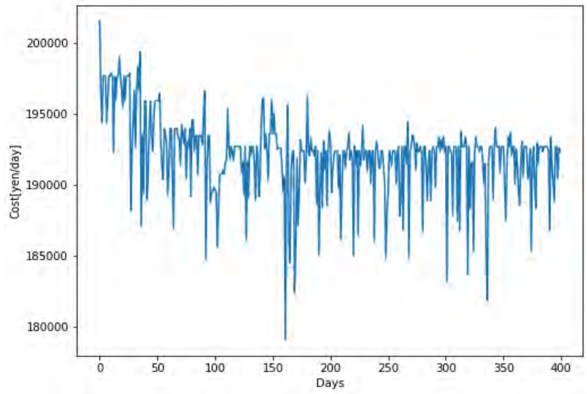

Figure 5. Total Cost Improvement by reinforcement learning.

At first, the cost is about 200,000 [yen/day]. However, as the aggregator spends more days, the aggregator's cost gets lower. And after about 150 days, the aggregator's cost becomes stable at about 192,000. This means the aggregator is able to learn the good strategy of controlling EV reward price. Compared to this system, if the aggregator doesn't rely on EVs and pays all the imbalance cost (baseline), daily cost is about 274,000 [yen /day]. Therefore, by relying on EVs, the aggregator can save 30% of its cost. Please note that we don't consider about initial cost, for example, the equipment cost to control households' EVs and tell the price to EV owners. This point is one of our future works.

To analyze the relationship between the imbalance cost and reward to EVs, we evaluate the daily reward supplied to EVs and the daily imbalance cost.

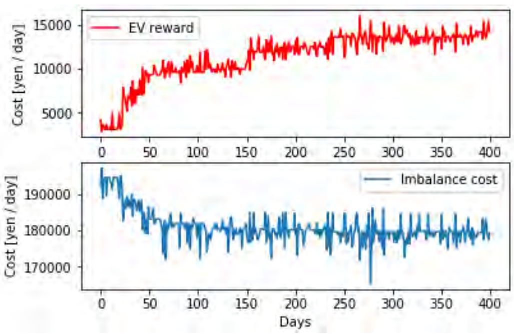

Figure 6. The relationship between the imbalance cost and EV reward.

Fig. 6 shows the relationship between the imbalance cost and the reward to EVs. The reward that the aggregator must pay to the EV owners increases as the aggregator learns. In contrast, the imbalance cost decreases as the aggregator learns. This means that the aggregator is able to learn how the aggregator should control the households' EVs to reduce imbalance penalty even the aggregator pays a lot of EV reward. From the perspective of the EV owners, they can get a reward from the aggregator by joining this schema. And this is also good for utility companies because they can operate their generation systems more efficiently with a less imbalance.

Next, to analyze how the imbalance penalty price affects the operation, we changed the value of the imbalance penalty (which is fixed to 24 [yen/kWh]) to 4, 8, 16, 24, and 32 [yen/kWh]. The result is shown in Fig. 7.

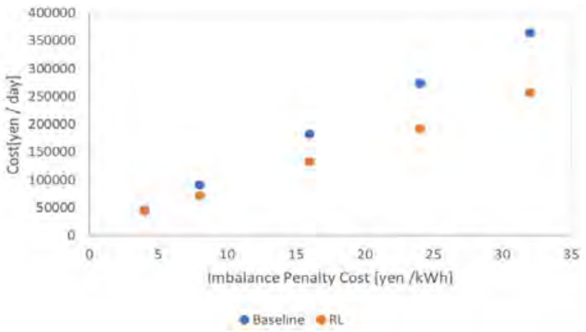

Figure 7. The cost analysis with different imbalance prices.

Fig. 7 shows that if the imbalance cost is too low, the aggregator can't learn well and decrease daily cost compared to the original case (No available EVs). However, if the price gets higher, the aggregator can save much more cost with EVs compared to the original case.

IV. Conclusion

This paper presented the aggregator's control strategy of households' EVs. By using reinforcement learning, the aggregator can find the good strategy without defining a well-defined optimization model. What the novelty we proposed was considering the future imbalance penalty cost in Japan. In the future, a lot of new PPS will have to consider the imbalance penalty cost that is derived from their renewables. In this paper, we proposed that the aggregator can reduce imbalance penalty by controlling households' EVs. And this is good for other stakeholders, such as EV owners and utility companies because EV owners can get a reward from the aggregator and the utility companies can operate their generation system efficiently.

Regarding our future work, as we describe at the result section, considering the equipment cost for the aggregator is one of future works. In addition, using other electricity resources such as battery and demand response (DR) is also one of the solutions to reduce the high imbalance penalty for the aggregators. Making the algorithm that can control not only EVs but also other electricity storage resources is also one of our future works.

References

[1] S. Vandael, B. Claessens, D. Ernst, T. Holvoet, and G. Deconinck, "Reinforcement learning of heuristic EV fleet charging in a day-ahead electricity market," *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 1795-1805, 2015.

[2] J. Tomić and W. Kempton, "Using fleets of electric-drive vehicles for grid support," *J. Power Sources*, vol. 168, no. 2, pp. 459-468, 2007.

[3] K. Takeda, M. Okazaki, G. T. Nhut, R. Matsuhashi, and R. Ariyoshi, "A study on using electric vehicles for load frequency control in power systems," in *Proc. International Conference and Exhibition of Grand Renewable*, 2018.

[4] J. Hu, H. Morais, T. Sousa, and M. Lind, "Electric vehicle fleet management in smart grids: A review of services, optimization and control aspects," *Renew. Sustain. Energy Rev.*, vol. 56, pp. 1207-1226, Apr. 2016.

[5] D. O'Neill, M. Levorato, A. Goldsmith, and U. Mitra, "Residential demand response using reinforcement learning," in *Proc. First IEEE International Conference on Smart Grid Communications*, 2010, pp. 409-414.

[6] I. Dusparic, C. Harris, A. Marinescu, V. Cahill, and S. Clarke, "Multi-agent residential demand response based on load forecasting," in *Proc. 1st IEEE Conference on Technologies for Sustainability (SusTech)*, 2013, pp. 90-96.

[7] T. Pinto, Z. Vale, F. Rodrigues, I. Praca, and H. Morais, "Cost dependent strategy for electricity markets bidding based on adaptive reinforcement learning," in *Proc. 16th International Conference on Intelligent System Applications to Power Systems*, 2011, pp. 1-6.

[8] A. L. Dimeas and N. D. Hatziargyriou, "Multi-agent reinforcement learning for microgrids," in *Proc. IEEE PES General Meeting*, 2010, pp. 1-8.

[9] W. Shi and V. W. S. Wong, "Real-time vehicle-to-grid control algorithm under price uncertainty," in *Proc. IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2011, pp. 261-266.

[10] F. Ruelens, B. J. Claessens, S. Vandael, S. Iacovella, P. Vingerhoets, and R. Belmans, "Demand response of a heterogeneous cluster of electric water heaters using batch reinforcement learning," in *Proc. Power Systems Computation Conference*, 2014, pp. 1-7.

[11] K. Valogianni, W. Ketter, and J. Collins, "Smart charging of electric vehicles using reinforcement learning," in *Proc. AAAI Workshop: Trading Agent Des.*, 2013, pp. 41-48.

[12] M. Stemmer, U. Wagner, and R. Matsuhashi, "Development of a model to estimate solar power generation output based on weather forecast information," *Japan Council for Renewable Energy*, no. 1, pp. 3-6, 2017.

[13] R. Matsuhashi, "Development of an optimization model to manage imbalances in power systems by local power retail companies," in *Proc. 7th International Conference on Power Science and Engineering*, 2018.

Ryuhei Furuta born in 1996, he earned the Bachelor of Engineering degree from the Department of Electrical and Electronic Engineering, the University of Tokyo in 2018. Currently, he is a master student at Department of Electrical Engineering and Information Systems, Graduate School of Engineering, The University of Tokyo. His research interests are energy economics, analysis of energy systems, and global environmental issues.

Ryuji Matsuhashi born in 1963, he earned the Bachelor of Engineering degree from the Department of Electronics, Faculty of Engineering, the University of Tokyo in 1985, and the Doctor of Engineering in 1990 from the same department. He became an Associate Professor at the Department of Geosystem Engineering, Faculty of Engineering, the University of Tokyo since 1994 after serving as the Research Associate of the same from 1990 to 1993. He has become an Associate Professor at the Institute of Environmental Studies, Graduate School of Frontier Sciences, The University of Tokyo, since 1999. Next, he has become a Professor at the same institute of the University of Tokyo, since 2003. Then he has

become a Professor at the Electrical Engineering and Information Systems, Graduate School of Engineering in the same university, since 2011. His intensive works in the analysis of energy systems and global environmental issues produced various books and papers. He is currently a member of Japan Society of Energy and Resources, the Institute of Electrical Engineers of Japan, the Japan Institute of Energy.