Deep Reinforcement Learning-based SOH-aware Battery Management for DER Aggregation

Shotaro Nonaka   Daichi Watari   Ittetsu Taniguchi   Takao Onoye
Graduate School of Information Science and Technology, Osaka University
1-5 Yamadaoka, Suita, Osaka, Japan
{nonaka.shotaro,watari.daichi,i-tanigu,onoye}@ist.osaka-u.ac.jp

ABSTRACT

In smart energy systems, batteries, which assume an important role in filling the temporal gap between generation and consumption, are expected to be a potential distributed energy resource (DER). A resource aggregator (RA) has emerged to collect various DERs to extract demand-side flexibility, and various methods have been proposed based on reinforcement learning. Since battery degradation is unavoidable during utilization, battery management is required to minimize it. This paper proposes state-of-health (SOH)-aware battery management based on deep reinforcement learning. Our experimental results demonstrate an average battery lifetime improvement of 11.2%.

CCS CONCEPTS

• **Hardware → Smart grid**; • **Computing methodologies → Planning under uncertainty**.

KEYWORDS

battery management, SOH (state-of-health), DER (distributed energy resources), aggregation

ACM Reference Format:
Shotaro Nonaka Daichi Watari Ittetsu Taniguchi Takao Onoye. 2022. Deep Reinforcement Learning-based SOH-aware Battery Management for DER Aggregation. In *Third ACM SIGEnergy Workshop on Reinforcement Learning for Energy Management in Buildings & Cities (RLEM) (RLEM '22), November 9–10, 2022, Boston, MA, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3563357.3566166

1 INTRODUCTION

Extracting the demand-side flexibility is a big motivation for the stabilization of the power grid, and it will be more required for utilizing renewable energy as much as possible. Behind-the-meter distributed energy resources (DERs) play important extraction roles, and resource aggregators (RAs) have emerged to concentrate various DERs to extract demand-side flexibility. Examples of flexibility services by RAs include ancillary services [12] and energy arbitrage [9].

Batteries are potential DERs for providing flexibility services and filling the temporal gap between generation and consumption [10]. However, batteries are unavoidably degraded during their use, and such degradation affects their lifetime. Thus controlling battery degradation is necessary.

Various research on RAs has been proposed, including the aggregated control of electric vehicles (EVs), heating, ventilation, and air conditioning (HVAC), etc. Yi et al. proposed an EV aggregation method based on mathematical programming [11]. Liu et al. proposed a battery aggregation method based on an optimal scheduling approach [4]. Iacovella et al. proposed a HVAC aggregation method using a three-step optimization scheme [3]. These researches demonstrated the efficiency of DER aggregation. However, these approaches require detailed models; based on reinforcement learning, model-free approaches have recently been proposed [2, 6, 8].

Qian et al. proposed an EV aggregation method to minimize the total driving time and the EV-charging cost [6]. Qian's method supports the different characteristics of EVs and also demonstrated the efficiency of deep reinforcement learning. Taboga et al. proposed an energy management framework, which covers RAs and multiple prosumers [8]. Taboga's method mainly targeted the HVAC of individual prosumers and achieved peak power shaving while simultaneously maintaining acceptable comfort levels. Sanchez et al. proposed a battery aggregation method based on deep reinforcement learning for grid stability [2]. Their method assumed that battery charge/discharge is managed by RA without addressing battery degradation.

This paper proposes a state-of-health (SOH)-aware battery management method based on deep reinforcement learning. Our proposed aggregation method realizes the battery aggregation to realize the flexibility requirement and to minimize the battery degradation based on Millner's SOH model [5]. Our experimental results demonstrate the effectiveness of our proposed method.

The rest of our paper is organized as follows. Section 2 explains the system model, and Section 3 formulates the problem as a Markov Decision Process (MDP) and solves it with a deep reinforcement learning algorithm. Experimental results are described in Section 4, and Section 5 concludes this paper.

2 SYSTEM MODEL

2.1 System Overview

Figure 1 represents a system overview of this research. The system model is composed of a balancing market, a resource aggregator (RA), and N prosumers. For the given requirements from the balancing market, the RA sends orders to all the prosumers, and the net demand is updated due to the battery charge/discharge changes

RLEM '22, November 9–10, 2022, Boston, MA, USA
© 2022 Association for Computing Machinery.
ACM ISBN 978-1-4503-9890-9/22/11…$15.00
https://doi.org/10.1145/3563357.3566166

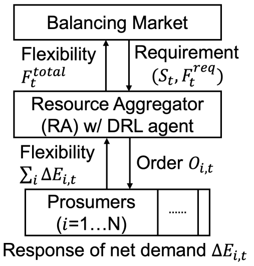

Figure 1: System Overview

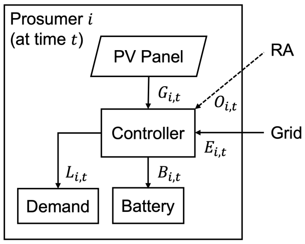

Figure 2: Prosumer Model

made by them. These responses are aggregated as flexibility, and requirements are satisfied when aggregation has succeeded. The research question of this paper investigates how to aggregate the flexibility considering the degradation of each prosumer's battery.

Let S_t, F_t^{req}, and $O_{i,t}$ be the required type of flexibility at time t, its required amount at time t, and the prosumer orders, respectively. Required type S_t takes 1 in upward demand cases, -1 in downward demand cases, and 0 when nothing is required. Required amount F_t^{req} represents the ratio over the maximum flexibility of the prosumers. Order $O_{i,t} \in [-1, 1]$ to prosumer i also represents the ratio over the maximum charge/discharge amount of prosumer i.

The prosumers respond by battery charge/discharge changes based on order $O_{i,t}$. When the change of the net demand of prosumer i is represented as $\Delta E_{i,t}$, the aggregated flexibility is represented by $\sum_i \Delta E_{i,t} = F_t^{total}$. Thus, the requirement is satisfied in case $|F_t^{total}| = F_t^{req}$.

2.2 Prosumer Model

Figure 2 shows a prosumer model, which is composed of a PV panel, demand from home appliances and a battery, and a controller that manages the power flow. Let $G_{i,t}$, $L_{i,t}$, $B_{i,t}$, and $E_{i,t}$ be the PV generation, the demand, the battery charge, and the purchased energy, respectively. Then the following equation must hold:

$$G_{i,t} + E_{i,t} = L_{i,t} + B_{i,t}. \tag{1}$$

Notice that $B_{i,t}$ and $E_{i,t}$ are bidirectional. The negative values of $B_{i,t}$ and $E_{i,t}$ represent the battery discharge and selling energy.

Let B_i^{cap}, $B_i^{max,ch}$, $B_i^{max,disch}$, and $SOC_{i,t}$ be the battery capacity, the maximum charge amount, the maximum discharge amount, and the state-of-charge of prosumer i's battery at time t, respectively. Then the following formulas must hold:

$$SOC_{i,t+1} = SOC_{i,t} + B_{i,t}/B_i^{cap}, \tag{2}$$

$$B_i^{max,disch} \le B_{i,t} \le B_i^{max,ch}. \tag{3}$$

When we denote $B_{i,t}^{r,ch}$ and $B_{i,t}^{r,disch}$ as the available charge/discharge amount at $SOC_{i,t}$, respectively, $B_{i,t}^{r,ch}$ and $B_{i,t}^{r,disch}$ are defined as follows:

$$B_{i,t}^{r,ch} = (SOC_i^{max} - SOC_{i,t}) \cdot B_i^{cap}, \tag{4}$$

$$B_{i,t}^{r,disch} = (SOC_{i,t} - SOC_i^{min}) \cdot (-B_i^{cap}), \tag{5}$$

where SOC_i^{max} and SOC_i^{min} are the upper- and lower-bounds of the SOC level.

For given $G_{i,t}$ and $L_{i,t}$, the controller decides battery charge amount $B_{i,t}$ based on the following rule:

$$B_{i,t} = \begin{cases} \min(G_{i,t} - L_{i,t}, B_{i,t}^{r,ch}, B_i^{max,ch}) & (G_{i,t} \ge L_{i,t}) \\ \max(G_{i,t} - L_{i,t}, B_{i,t}^{r,disch}, B_i^{max,disch}) & (\text{otherwise}). \end{cases} \tag{6}$$

Then purchased power $E_{i,t}$ is calculated by Eq. 1.

2.3 Flexibility Extraction

After receiving order $O_{i,t} \in [-1, 1]$ from the RA, prosumers respond by changing their battery charge/discharge amount. If $O_{i,t} = 1$, they increase their demand by an additional full charge. If $O_{i,t} = -0.5$, they decrease their demand by an additional half discharge. However, since the prosumers determined their battery charge/discharge amount by Eq. 6, such received orders might fail to be satisfied due to their battery states. For example, if a battery has already been fully charged ($SOC_{i,t} = 1$), an additional full charge due to $O_{i,t} = 1$ is impossible. If a battery plans to fully discharge by Eq. 6, an additional discharge due to $O_{i,t} = -1$ is also impossible. In such cases, the required flexibility cannot be satisfied due to the obvious physical constraints. Therefore, the change of net demand $\Delta E_{i,t}$ is decided as follows:

$$\Delta E_{i,t} = \begin{cases} \min(O_{i,t} \cdot F_i^{max}, B_{i,t}^{r,ch}, B^{avl,c}) & (O_{i,t} \ge 0) \\ \max(O_{i,t} \cdot F_i^{max}, B_{i,t}^{r,disch}, B^{avl,d}) & (\text{otherwise}) \end{cases} \tag{7}$$

where $B^{avl,c}$ and $B^{avl,d}$ are defined as follows:

$$B^{avl,c} = B_i^{max,ch} - B_{i,t}, \tag{8}$$

$$B^{avl,d} = -B_i^{max,disch} - B_{i,t}. \tag{9}$$

The change of net demand $\Delta E_{i,t}$ is aggregated at the RA. The aggregation succeeded in case $F_t^{total} = \sum_i \Delta E_{i,t}$ and satisfies the requirements of the balancing market.

2.4 Battery Degradation Model

Battery degradation is defined as the phenomenon of the decrease of battery capacity after utilization. This section briefly introduces Millner's battery degradation model [5]. Let X_{full}^{init} and X_{full} be the initial (new) capacity and the current capacity. The ratio of the degradation represented by SOH is defined as follows:

$$SOH = \frac{X_{full}}{X_{full}^{init}}. \tag{10}$$

Millner estimated battery degradation with the following degradation model. Let $L_{cycle,m}$ be the degradation ratios after m charge/discharge iteration. SOH after M charge/discharge iteration is defined as follows:

$$SOH = 1 - \sum_{m}^{M} L_{cycle,m}. \tag{11}$$

Let SOC^{swing} and SOC^{ave} be the swing at the charge/discharge cycle and the average SOC level. Then $L_{cycle,m}$ is defined as follows:

$$L_{cycle,m} = L_2 \cdot \exp\left[K_T(T_B - T_{ref}) \cdot \frac{T_{ref} + 273}{T_B + 273}\right], \tag{12}$$

$$L_2 = L_1 \cdot \exp\left[4K_{SOC}(SOC^{avg} - 0.5)\right] \cdot \left(1 - \sum_{m}^{M-1} L_{cycle,m}\right), \tag{13}$$

$$L_1 = K_{CO} \cdot \exp\left[(SOC^{swing} - 1)\frac{T_{ref} + 273}{K_{EX}(T_B + 273)}\right] + 0.2\frac{\tau}{\tau_{life}}, \quad (14)$$

where K_{CO}, K_{EX}, K_{SOC}, and K_T are the battery specific parameters. T_{ref} and T_B are the reference and battery temperatures. τ represents the time in seconds of the m-th charge/discharge cycle, and τ_{life} denotes the total expected calendar life in seconds. Due to space limitations, we omit a detailed explanation of the model.

The degradation model includes various physical parameters. On the other hand, from an operation viewpoint, we still have potential knobs SOC^{swing} and SOC^{ave} to control the battery degradation. This paper's idea is to control battery degradation through these knobs during battery aggregation.

3 PROBLEM FORMULATION

We formulate the problem as a Markov Decision Process (MDP) to handle it with a deep reinforcement learning algorithm. An MDP consists of a set comprised of a state, an action, and a reward for each time step t. The agent decides action $\mathbf{a}_t$ based on system state $\mathbf{s}_t$ and can observe new state $\mathbf{s}_{t+1}$ and reward R_t.

System state $\mathbf{s}_t$ is defined as follows:

$$\mathbf{s}_t = \left(\mathbf{SOC}_t, S_t, F_t^{req}, t\right), \quad (15)$$

where $\mathbf{SOC}_t$ is denoted as follows:

$$\mathbf{SOC}_t = \left(SOC_{1,t}, ..., SOC_{N,t}\right). \quad (16)$$

Action $\mathbf{a}_t$ is defined as the vector of the orders as follows:

$$\mathbf{a}_t = \left(O_{1,t}, ..., O_{N,t}\right). \quad (17)$$

The reward design is critical to efficiently train the agent. The objectives of this problem are to extract the required flexibility and minimize the battery degradation. Thus we introduce two reward terms: r_t^{create} and r_t^{deg}. The former denotes the reward by flexibility extraction, and the latter denotes it by degradation control. We define reward R_t as follows:

$$R_t = \omega_1 \cdot r_t^{create} - \omega_2 \cdot r_t^{deg}, \quad (18)$$

where ω_1 and ω_2 are weight parameters that range from 0 to 1.

When we define $F_t^{dif} = F_t^{req} - |F_t^{total}|$, reward term r_t^{create} is defined as follows:

$$r_t^{create} = \begin{cases} -(F_t^{dif})^2 & (1 + \delta \geq |F_t^{total}/F_t^{req}| \geq 1 - \delta) \\ -P \cdot (F_t^{dif})^2 & (\text{otherwise}), \end{cases} \quad (19)$$

where δ and P are the acceptable error and the penalty. In this model, flexibility extraction succeeded if $1 + \delta \geq |F_t^{total}/F_t^{req}| \geq 1 - \delta$. Otherwise, flexibility extraction failed, and a penalty is imposed.

Reward term r_t^{deg} is defined as follows:

$$r_t^{deg} = \sum_{i=1}^{N} \exp\left[4K_{SOC}(SOC_i^{ave} - 0.5)\right], \quad (20)$$

where K_{SOC} and SOC_i^{ave} are the battery parameters and the average SOC level of prosumer i's battery. The reward term by battery degradation is based on the battery degradation model, especially Eq. 13. Battery degradation is affected by the battery usage pattern, especially the average level and the fluctuation, as explained in Section 2.4. Therefore, the reward term from the battery degradation control includes the part of the degradation model affected by battery usage.

Table 1: Overview of Experimental Results

	Success	MAE	MAPE	Lifetime [days]		
Method	Rate [%]	[kW]	[%]	Min	Ave	Max
w/o r_{deg}	63.6	1.77	12.5	1,889	6,903	10,781
w/ r_{deg}	58.2	2.49	17.6	4,925	7,679	10,990

4 EXPERIMENTAL RESULTS

We implemented our proposed algorithm in python to demonstrate the efficiency of our proposed method. We trained the RA agent using Proximal Policy Optimization [7]. The total training step is 3×10^6 steps, the learning rate is 2.5×10^{-4}, the discount rate is 0.85, the clip range is 0.1, and the batch size is 256. For the reward function in Eq. 18, the weight parameters of ω_1 and ω_2 are set to 1.

In this experiment, we assumed ten prosumers ($N = 10$), each of whom has a battery with 11.2kWh capacity ($B_i^{cap} = 11.2$). The time resolution is five minutes, and each episode (one day) consists of 288 steps. We assume that a requirement is received every 30 minutes, each of which is announced five minutes before the flexibility extraction begins. The requirements were randomly generated from every possible combination of $S_t = \{-1, 0, 1\}$ and $F_t^{req} = \{11.2, 14.0, 16.8\}$. The flexibility is extracted in 10-minute durations.

Based on the received requests, the RA sends orders $O_{i,t}$ to all the prosumers. In this experiment, we set $O_{i,t} \in [-0.5, 0.5]$ due to the efficiency of learning, the convenience of our preferred prosumers, etc. We also set $\delta = 0.1$, and flexibility extraction succeeded in case $1.1 \geq |F_t^{total}/F_t^{req}| \geq 0.9$. The battery parameters were from the data sheet of Panasonic Li-ion battery LJB1156. The data set was from the UMass Smart* Data Set [1].

In this experiment, the battery depleted its lifetime when SOH reached 0.8, and the simulation was iterated until all the SOHs of all the batterys reached 0.8. Our proposed method was evaluated by both its aggregation results and battery lifetimes. Table 1 summarizes the experimental results. "w/o r_{deg}" denotes the case where the reward function does not include term r_{deg}. "w/ r_{deg}" denotes our proposed method, which is when the reward function includes term r_{deg}. The success rate represents the ratio to satisfy $1 + \delta \geq |F_t^{total}/F_t^{req}| \geq 1 - \delta$. MAE and MAPE were calculated between required flexibility F_t^{req} and extracted flexibility F_t^{total}. As shown in Table 1, although the aggregation results worsened, the battery lifetimes basically improved. For example, the success rate fell by 5.4%, and MAPE increased by 5.1%. However, the average lifetimes improved by 11.2%, especially the minimum lifetime, which drastically improved by 2.6 times.

Figure 3 shows the battery lifetime distribution. X-axis represents battery IDs, and Y-axis represents their lifetimes. The lifetimes of batteries #4, #7, and #8 were very short in case "w/o r_{deg}". However, the proposed method largely extended the lifetimes of these weak batteries by 2-3 times. Thus, the average lifetime also rose by more than 10%.

Figures 4 and 5 show the SOC traces (the first three days) of batteries #7 and #8. The average SOC level of "w/ r_{deg}" was obviously

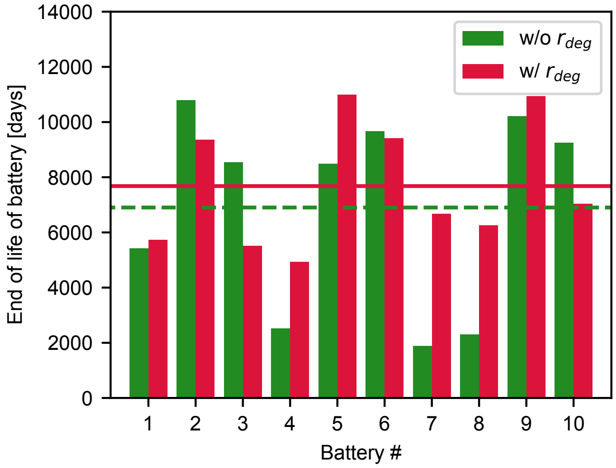

Figure 3: Battery Lifetime Distribution

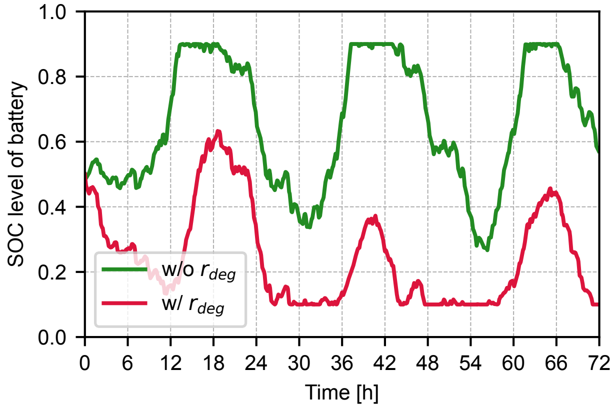

Figure 4: SOC Trace of Battery #7

lower than that of "w/o r_{deg}" for both cases. Our proposed battery aggregation includes a reward term for the battery degradation model shown in Eq. 20. Thus the average SOC level fell, and we expect battery degradation to improve with our proposed method.

5 CONCLUSION

This paper proposed a deep reinforcement learning-based battery aggregation method by considering battery degradation. We introduced Millner's SOH model to control the unavoidable battery degradation. The battery aggregation problem was formulated as a Markov Decision Process (MDP) and solved using deep reinforcement learning. Experimental results demonstrated the efficiency of the proposed method. The average battery lifetime was improved by 11.2%. Our proposed method largely extended the lifetimes of weak batteries 2-3 times.

This is the first paper that addressed both aggregation and battery degradation control. Our proposed method improved the battery

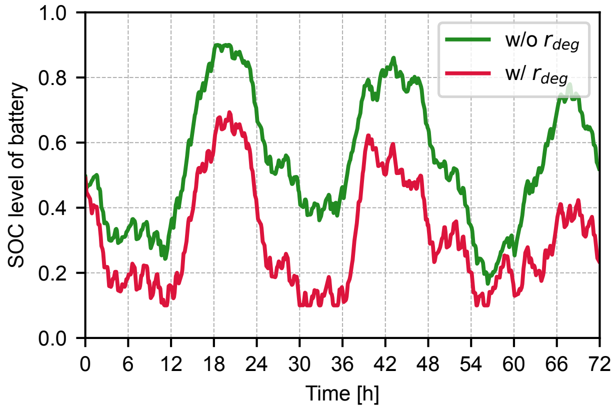

Figure 5: SOC Trace of Battery #8

degradation, but the performance of the aggregation should be improved more. Thus, improving aggregation is critical future work.

ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI Grant Numbers JP22H03697, JP21J10312, and DAIKIN Industries, Ltd.

REFERENCES

[1] UMass Smart* Data Set for Sustainability. https://traces.cs.umass.edu/index.php/Smart/Smart. [Accessed;2022-08-15].
[2] Gorostiza, F. S., and Gonzalez-Longatt, F. M. Deep reinforcement learning-based controller for SOC management of multi-electrical energy storage system. *IEEE Transactions on Smart Grid 11*, 6 (11 2020), 5039–5050.
[3] Iacovella, S., Ruelens, F., Vingerhoets, P., Claessens, B., and Deconinck, G. Cluster control of heterogeneous thermostatically controlled loads using tracer devices. *IEEE Transactions on Smart Grid 8*, 2 (Mar. 2017), 528–536.
[4] Liu, K., Chen, Q., Kang, C., Su, W., and Zhong, G. Optimal operation strategy for distributed battery aggregator providing energy and ancillary services. *Journal of Modern Power Systems and Clean Energy 6*, 4 (7 2018), 722–732.
[5] Millner, A. Modeling lithium ion battery degradation in electric vehicles. In *Proc. of 2010 IEEE Conference on Innovative Technologies for an Efficient and Reliable Electricity Supply (CITRES)* (9 2010), pp. 349–356.
[6] Qian, T., Shao, C., Wang, X., and Shahidehpour, M. Deep reinforcement learning for EV charging navigation by coordinating smart grid and intelligent transportation system. *IEEE Transactions on Smart Grid 11*, 2 (3 2020), 1714–1723.
[7] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., and Klimov, O. Proximal policy optimization algorithms.
[8] Taboga, V., Bellahsen, A., and Dagdougui, H. Deep reinforcement learning for peak load reduction in aggregated residential houses. In *Proc. of 2020 IEEE Power Energy Society General Meeting (PESGM)* (8 2020), pp. 1–5.
[9] Terlouw, T., AlSkaif, T., Bauer, C., and van Sark, W. Multi-objective optimization of energy arbitrage in community energy storage systems using different battery technologies. *Appl. Energy 239* (Apr. 2019), 356–372.
[10] Watari, D., Taniguchi, I., Goverde, H., Manganiello, P., Shirazi, E., Catthoor, F., and Onoye, T. Multi-time scale energy management framework for smart PV systems mixing fast and slow dynamics. *Applied Energy 289* (May 2021), 116671.
[11] Yi, Z., Xu, Y., Gu, W., and Wu, W. A multi-time-scale economic scheduling strategy for virtual power plant based on deferrable loads aggregation and disaggregation. *IEEE Transactions on Sustainable Energy 11*, 3 (7 2020), 1332–1346.
[12] Zhu, D., and Zhang, Y.-J. A. Optimal coordinated control of multiple battery energy storage systems for primary frequency regulation. *IEEE Trans. Power Syst. 34*, 1 (Jan. 2019), 555–565.