Improving duck curve by dynamic pricing and battery scheduling based on a deep reinforcement learning approach

Daichi Watari, Ittetsu Taniguchi, Takao Onoye
Graduate School of Information Scienece and Technology, Osaka University, Japan
{watari.daichi,i-tanigu,onoye}@ist.osaka-u.ac.jp

ABSTRACT

The duck curve is becoming a worldwide problem due to the rapid introduction of photovoltaic systems. A resource aggregator (RA) has emerged to provide flexible solutions through demand response and aggregating prosumers. This paper proposes a deep reinforcement learning based strategy of the RA that dispatches dynamic pricing to the prosumers and schedules its battery system to improve the duck curve. The results show that appropriate reward functions can improve the standard deviation and peak-to-average ratio of netload by up to 51.6% and 14.8%, respectively.

CCS CONCEPTS

• **Hardware** → **Smart grid**; • **Computing methodologies** → **Reinforcement learning**.

KEYWORDS

Duck curve, Dynamic pricing, Battery, Deep reinforcement learning

ACM Reference Format:
Daichi Watari, Ittetsu Taniguchi, and Takao Onoye. 2021. Improving duck curve by dynamic pricing and battery scheduling based on a deep reinforcement learning approach. In *The 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '21), November 17–18, 2021, Coimbra, Portugal.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3486611.3492232

1 INTRODUCTION

The high penetration of solar energy to prosumers causes serious problems such as the duck curve [1]. The duck curve is a graph of total netload change that shows the great imbalance between peak demand and solar generation. In recent years, a resource aggregator (RA) has an important role to coordinate prosumers' demand and dispatch dynamic pricing programs as the demand response.

Previous research has been conducted on the dynamic pricing and the aggregator's strategy. A model-based optimization approach calculates the optimal retail prices based on demand-supply balances [2]. However, it generally make the impractical assumption of having complete knowledge about the prosumers, and the computational cost is expensive. Recently, a reinforcement learning (RL) and a deep RL (DRL), which are model-free approaches, have succeeded to solve a complex problem of power systems. Lu et al. have proposed a model-free RL-based dynamic pricing method on an electricity retailer to maximize the profit for targeting the conventional consumer [4]. To the best of our knowledge, there is no significant study on a model-free method to improve the duck curve by dynamic pricing and battery scheduling.

BuildSys '21, November 17–18, 2021, Coimbra, Portugal

ACM ISBN 978-1-4503-9114-6/21/11.
https://doi.org/10.1145/3486611.3492232

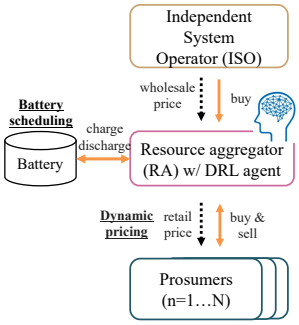

Figure 1: System overview

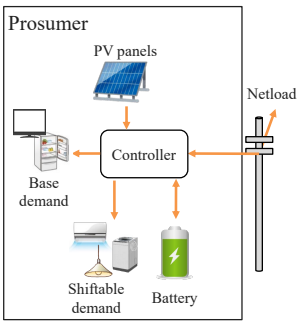

Figure 2: Prosumer model

In this paper, we propose a model-free DRL-based strategy for the RA to improve the duck curve. The proposed strategy optimizes both the retail prices for each prosumer in the dynamic pricing program and the RA's battery operation. The contributions of this paper are: (1) a DRL-based strategy is developed to optimize the dynamic pricing and battery scheduling simultaneously and (2) a design of the reward function is carefully explored to improve the duck curve. Our poster shows the details and the effectiveness of the proposed method.

2 PROBLEM SETTING

We target a hierarchical electricity market model composed of an independent system operator (ISO), a resource aggregator (RA), and prosumers, as shown in Fig 1. The RA aggregates prosumer's demand and joins a wholesale electricity market organized by the ISO. The RA dispatches a dynamic pricing program to the prosumers and sells/buys the electricity at time-varying retail prices. Furthermore, the RA has a large-capacity battery system, and it can be charged/discharged to increase/reduce the netload. On the other hand, the prosumers are equipped with a photovoltaic (PV) panel, a battery, base demand, and shiftable demand, as illustrated in Fig 2. The behavior of the prosumers, i.e., the operation of the shiftable demand and the battery, is expected to change responding to the retail price. Note that we assume that the prosumers are myopic and consider the current price only in the same way as [4]. Thus, if the current retail price is high, the prosumers reduce the shiftable demand and discharge their battery, vice versa.

This study focuses on the RA's decision-making for DP-BS (Dynamic Pricing and Battery Scheduling) problem. The ISO first informs the RA of the wholesale price, and the prosumers report the

Table 1: Reward terms for improving duck curve

Reward	Description
$r^{duck}_{t,dev} = (E^{net}_t - e^{net,avg}_d)^2$	Quadratic penalty of deviation from average
$r^{duck}_{t,diff} = (E^{net}_t - E^{net}_{t-1})^2$	Quadratic penalty of time diff. of total netload
$r^{duck}_{t,quad} = (E^{net}_t)^2$	Quadratic penalty of total netload
$r^{duck}_{t,cubic} = \mathrm{sign}(E^{net}_t) \cdot (E^{net}_t)^3$	Cubic penalty of total netload
$r^{duck}_{t,no} = 0$	No reward

expected netload to the RA. The RA determines the charge/discharge amount of the RA's battery and the retail prices for each prosumer, after that, informing the prosumers of the retail price. The prosumers decide the operation of their battery and shiftable demand based on the retail price and report the actual netload to the RA. Finally, the RA calculates the total netload, denoted by E^{net}_t, and trades it with the ISO. The RA learns appropriate DP-BS strategy from the above interactions with the ISO and the prosumers. Through the DP-BS, the RA aims to maximize social welfare, including the duck curve improvement.

3 METHODOLOGY

We formulates the DP-BS problem as the Markov Decision Process (MDP) to handle the problem by the DRL algorithm. The MDP mainly consists of a set of state, action, and reward for each time step t. The agent decides the action a_t based on the system state s_t, and then, can observe the new state s_{t+1} and reward R_t.

State. We assume that the observed state consists of a time index, the wholesale prices from the ISO, the pre-announced netload of each prosumers, the state-of-charge (SOC) of the batteries of prosumers and the RA.

Actions. The actions taken by the RA are the retail price for each prosumer and the operation of the RA's battery. The action spaces are assumed to be continuous, and the range of the retailed prices and the battery capacity is constrained by the upper/lower bounds.

Reward design. The designing of an appropriate reward function is critical to train the agent efficiently. The objectives of the DP-BS are to improve the RA's profit, the prosumer's cost, and the duck curve improvement. We assume that the reward function is set by some authoritative entity such as a utility:

$$R_t = \omega_1 \cdot P^{ra}_t - \omega_2 \cdot C^{pro}_t - (1 - \omega_1 - \omega_2) \cdot r^{duck}_t + P(\lambda_{t,n}) \quad (1)$$

where ω_1 and ω_2 are the weight parameters with the range from 0 to 1. P^{ra}_t is the profit of the RA by selling the electricity to the prosumers. C^{pro}_t means the total electricity cost of the prosumers. r^{duck}_t is a reward term to improve the duck curve. $P(\lambda_{t,n})$ is a penalty function of the retail prices, which takes a high value as the retail price approaches the upper/lower bounds. This will prevent taking the extreme high/low prices. Note that these reward terms are normalized so that the mean is zero and the standard deviation is one to handle them equally. The average netload of the day d, denoted by $e^{net,avg}_d$, is assumed to be predicted before the start of the day. In this study, we define five different reward terms as r^{duck}_t as shown in Table 1. We will compare their performance in Sec. 4.

DRL algorithm. We train the RA agent using widely used Soft Actor Critic (SAC) [3]. The SAC is an off-policy actor-critic DRL method to learn a stochastic policy maximizing an entropy. We choose the SAC because it has achieved the state-of-the-art performance for continuous control tasks.

Table 2: Results of minimum and maximum netload and average standard deviation and PAR for each day's results with test dataset (1st Aug. - 7th Aug.) and different reward.

Reward	r^{duck}_{dev}	r^{duck}_{diff}	r^{duck}_{quad}	r^{duck}_{cubic}	r^{duck}_{no}
Min netload [kWh]	70.39	36.65	61.76	63.26	39.44
Max netload [kWh]	235.28	232.48	241.10	231.63	275.42
Avg. std of netload [kWh]	19.47	32.27	28.46	24.46	40.22
Avg. PAR	1.30	1.35	1.37	1.33	1.52

4 EXPERIMENTAL RESULTS

We implemented the simulator and the proposed algorithm in `python`. The time resolution of the actions, i.e., the intervals of DP-BS, were set to 30 min. We used the open-datasets for power consumption [5] and PV generation [6]. We also used the wholesale electricity prices downloaded from the California ISO. The periods of the datasets are: the training set is from 1st to 31st July 2017; the test set is from 1st to 7th August 2017. The RA has a 300kWh battery system and its SOC range is from 20% to 90%.The range of the retail prices is 1.5 times the min/max wholesale price. Both weights of ω_1 and ω_2 were set to 0.2. We assumed that there are ten prosumers, and they have a battery of 20, 30, or 40 kWh.

We evaluated what reward terms r^{duck}_t are most effective in improving the duck curve. The evaluation metrics for improving the duck curve are the average of the standard deviation and PAR (Peak to Average Ratio) of the netload for each day, in addition to the minimum and maximum netload. Table 2 shows the results for one week using the model well trained with one month of training data. From Table 2, we can see that the proposed method achieves the best standard deviation and PAR when $r^{duck}_{t,dev}$ is set, which can improve 51.6% and 14.8%, respectively, compared to the case with $r^{duck}_{t,no}$. The main reason for that is to increase the minimum netload by minimizing the deviation from the average netload.

We conclude that the proposed method can improve the duck curve in terms of both the standard deviation and the PAR. Future work includes developing the multi-agent RL integrating prosumer's control for further improving the duck curve.

ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI Grant No. JP21J10312.

REFERENCES

[1] P. Denholm, M. O'Connell, G. Brinkman, and J. Jorgenson. 2015. *Overgeneration from solar energy in california. a field guide to the duck chart.* Technical Report. National Renewable Energy Lab. (NREL).
[2] J. Ferdous, M. P. Mollah, M. A. Razzaque, M. M. Hassan, A. Alamri, G. Fortino, et al. 2017. Optimal dynamic pricing for trading-off user utility and operator profit in smart grid. *IEEE Trans. Syst. Man Cybern.* 50, 2 (2017), 455–467.
[3] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine. 2018. Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor. In *Proc. ICML*, Vol. 80. 1861–1870.
[4] R. Lu, S. H. Hong, and X. Zhang. 2018. A Dynamic pricing demand response algorithm for smart grid: Reinforcement learning approach. *Appl. Energy* 220 (2018), 220–230.
[5] C. Miller, A. Kathirgamanathan, B. Picchetti, P. Arjunan, J. Y. Park, Z. Nagy, et al. 2020. The Building Data Genome Project 2, energy meter data from the ASHRAE Great Energy Predictor III competition. *Sci Data* 7, 1 (2020), 368.
[6] the California Solar Initiative (CSI). [n.d.]. California Distributed Generation Statistics. https://www.californiadgstats.ca.gov/downloads/. Accessed: 2021-7-19.