Optimal Control of Battery System by Reinforcement Learning Considering Profitability

Takuya Goto
Degree programs in Systems and Information Engineering
University of Tsukuba
Tsukuba 305-8573, Ibaraki, Japan
s2220845@s.tsukuba.ac.jp

Daisuke Kodaira*
Institute of Systems and Information Engineering
University of Tsukuba
Tsukuba 305-8573, Ibaraki, Japan
daisuke.kodaira03@gmail.com
*Corresponding author

2023 International Conference on Power and Renewable Energy Engineering (PREE) | 979-8-3503-2190-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/PREE57903.2023.10370554

Abstract— **Government-led pilot experiments for Virtual Power Plants (VPPs) are being conducted in Japan. However, these experiments have faced challenges such as deviations between predicted and actual values of photovoltaic (PV) generation, as well as insufficient battery capacity, leading to imbalances and associated penalties that reduce the revenue of VPP aggregators. In this study, we propose a method that combines probabilistic PV forecasting and deep reinforcement learning to determine the charging and discharging schedules of batteries within the VPP. Our approach considers the risks associated with deviations between forecasted and actual values, enabling the generation of operational plans while considering these uncertainties.**

Keywords-component; virtual power plant; distributed energy resources; deep reinforcement learning; battery scheduling; PV generation forecast; energy trading

I. Introduction

Virtual Power Plants (VPPs) are attracting attention to enable the economic utilization of energy across the entire power system by aggregating distributed energy resources owned by consumers, such as solar photovoltaic (PV) systems and batteries. In Japan, it is envisioned that aggregators will remotely control these devices, aggregating power and engaging in trading on the spot market and day-ahead market. While pilot experiments are underway to promote VPP activation, challenges arise from imbalances caused by control failures, such as deviations between predicted and actual PV generation and insufficient battery capacity [1]. Imbalances refer to the difference between planned and actual traded energy quantities, which result in penalties imposed by the general transmission.

Several studies have addressed the scheduling of electric vehicles (EVs) and battery systems [2]–[6]. [2] proposes a linear programming-based battery scheduling approach that resolves the issue of simultaneous charging and discharging by introducing optimization variables. However, this method may introduce complexity and challenges associated with introducing specific variables. Additionally, the scheduling of many EVs and batteries is known to be NP-hard [3], making linear programming infeasible. Deep reinforcement learning (RL) is considered promising for addressing the NP-hardness and uncertainties in EV and PV demand and generation [4]. [6] discusses the exploration of optimal algorithms for battery control using deep reinforcement learning. Among the algorithms examined in the literature, such as Advantage Actor-Critic (A2C) and Double Deep Q Networks (DDQN) [5], Proximal Policy Optimization (PPO) is highlighted as an approach that maximizes profits and enables safe planning. While the authors propose LSTM (Long Short-Term Memory) -based PV output forecasting and PPO-based battery scheduling for profit optimization in an environment combining PV and battery systems, their PV forecasting is deterministic and does not consider the uncertainty of PV generation.

This study proposes a methodology that probabilistically forecasts PV generation and uses deep reinforcement learning to determine battery scheduling, considering supply-demand planning and imbalances in actual supply-demand. The proposed approach aims to reduce imbalance penalties, improve revenue, and verify the feasibility of the scheduled charge and/or discharge.

II. Methodology

A. System configuration

The system of this study is designed to operate the experimental equipment installed at the National Institute of Advanced Industrial Science and Technology (AIST) in Tsukuba City (Fig. 1). The program is developed based on the system configuration shown in Fig. 2, assuming a rated power of 2.0 kW for the PV system and a rated capacity of 4.2 kWh for the battery. The battery is assumed to charge solely from the power supplied by the PV system, independent of power demand or the status of other devices, without charging from the grid. Furthermore, all power supplied by the PV system and the battery is sent to the power grid, allowing for direct sale of the power generated by the PV system. The devices are listed in Table. 1.

Table. 1 Products Number

Products	Manufacturer	Model number
PCS*	OMRON	KPAC-B25
PCS (for PV)		KP40K2
Battery		KP-BU42-A

*PCS: Power Conditioning System

This work was supported by JSPS KAKENHI Grant Number JP23K13306.

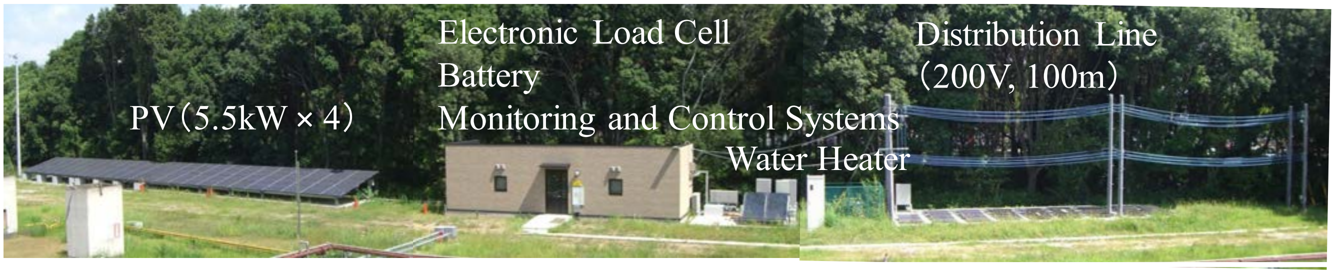

Fig. 1 Panoramic view of the experimental facility (AIST, Tsukuba City)

B. *Trading Flow*

The trading flow assumes a spot market where transactions take place on the actual day before supply and demand. This is illustrated in Fig. 3. In this system, since all the power generated by the PV system is sold, the consumer receives compensation from the distribution utility company for the amount of power generated. However, they need to pay an imbalance fee, which is the difference between the planned and actual amount of energy sold.

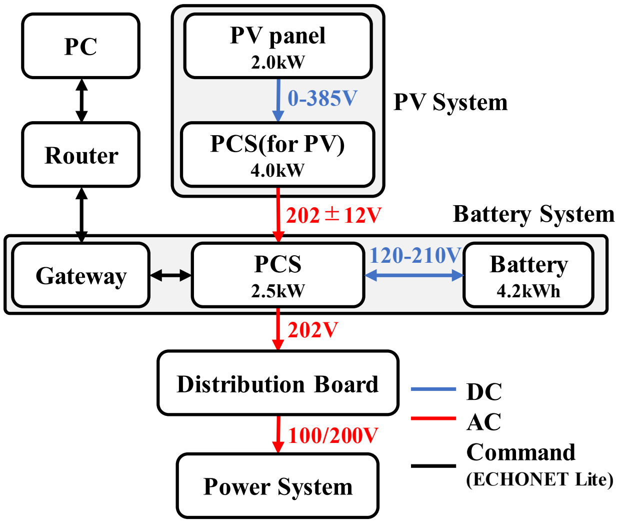

Fig. 2 System configuration

C. *Program model*

In this study, the program assumes power transactions on the day following the execution of the simulation. The program's structure is illustrated in Fig. 1. It consists of two main components: the prediction component for PV output and energy prices, and the planning component for battery charging and discharging. These components are integrated and executed together.

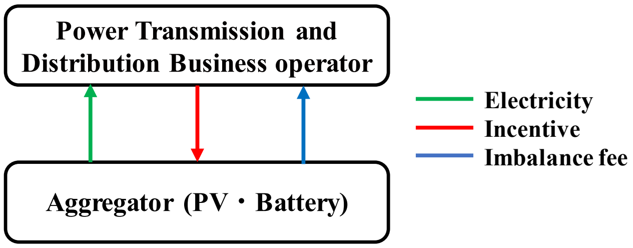

Fig. 3 Trading Flow

A common set of training data is prepared, which includes past PV output, energy prices, and actual weather measurements. After training on this data, the program obtains the predicted weather data for the next day. Using this data, the prediction component generates the required PV output and energy price predictions for the planning phase. These predictions are then input into the planning component, which outputs the charge/discharge plan for the battery on the following day. The plan is then sent as instructions to the actual equipment. The details of each prediction and planning component will be described in the following sections.

1) *PV generation prediction*

According to reference [7], it has been demonstrated that extracting highly weighted features and using them for prediction leads to higher accuracy in PV output forecasting. Therefore, in this study, a similar approach as in [7] is employed. During the training phase, four important features for prediction are selected, and the learning process is conducted using these features. After training, the PV output prediction is output in the form of upper and lower limits of a 95% prediction interval based on a normal distribution. The training is performed using actual PV output and weather measurements.

2) *Price Prediction*

The prediction of energy transaction prices and imbalance fees is conducted using LSTM models. However, for the purpose of operating the entire system in this study, the focus is not on achieving high accuracy. Unlike in scenario 1), the results are deterministic without prediction intervals. The training is performed using actual PV output and weather measurements.

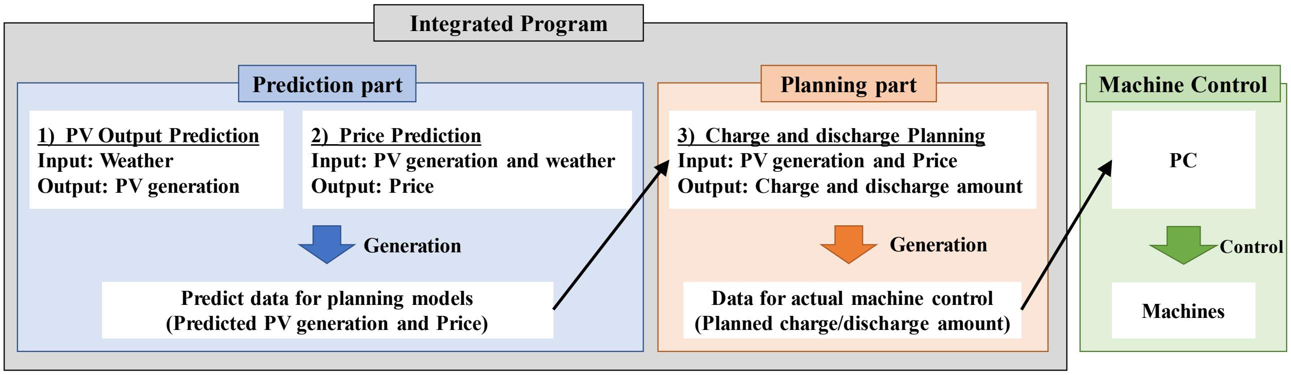
Integrated Program
Prediction part
1) PV Output Prediction
Input: Weather
Output: PV generation
2) Price Prediction
Input: PV generation and weather
Output: Price
Generation
Predict data for planning models
(Predicted PV generation and Price)
Planning part
3) Charge and discharge Planning
Input: PV generation and Price
Output: Charge and discharge amount
Generation
Data for actual machine control
(Planned charge/discharge amount)
Machine Control
PC
Control
Machines

Fig. 4 Program Configuration

3) *Charge and discharge Planning*

The charge/discharge plan for the battery is determined using deep reinforcement learning. The goal of this study is to generate a battery operating model that maximizes profitability. To achieve this, unrealistic operations such as charging beyond the rated capacity of the battery or discharging more than what is charged are considered infeasible actions. Rewards are defined during the selling of energy. By learning to maximize these rewards, the aim is to maximize profitability. The setting of rewards is described as follows (1)-(9):

$$\begin{cases} E_{in,k} = 0 \\ E_{out,k} = a_k \end{cases} (act_k > 0)$$
$$\begin{cases} E_{in,k} = a_k \\ E_{out,k} = 0 \end{cases} (act_k < 0) \qquad (1)$$

where act_k is the action value determined by deep reinforcement learning. $E_{in,k}$ is the amount of charging power [kW]. $E_{out,k}$ is the amount of discharging power [kW].k is the time index $(0 \leq k < 48)$. In (1), act_k is the output of discharge at positive values and the input of charge at negative values.

$$R_{1,k} = \begin{cases} E_{in,k} \times p_k(PV_k < -E_{in}) \\ (E_{c,k} - E_{out,k} \times T) \times p_k(E_{c,k} < E_{out,k} \times T) \end{cases} \qquad (2)$$

where p_k is the transaction price per 30 minutes [JPY/kWh/30min]. PV_k is PV output prediction [kW]. $E_{c,k}$ is the amount of energy stored by the battery [kWh]. T is the duration time (0.5h). In (2), R_1 represents the penalty incurred in the event of a charge/discharge operation that cannot be realized; R_1 is negative value and should ideally be zero.

$$R_{2,k} = (-E_{c,k}) \times p_k\ (SoC_k > 100\%) \qquad (3)$$

where SoC_k is the remaining charge of the battery [%]. In (3), R_2 represents the penalty to be given if the SoC exceeds 100%; R_2 is negative and should ideally be zero. Since (1) already penalizes discharging more than the remaining battery capacity, so there is no specific penalty defined for SOC below 0%.

$$R_{3,k} = E_{out,k} \times p_k(act_k > 0\,, E_{c,k} \geq E_{out,k} \times T) \qquad (4)$$

In (4), R_3represents the actual profit obtained solely from discharging the battery, and it is desirable for this value to be higher. From the above, the total reward obtained in one step is (5):

$$R_k = \sum_{n=0}^{t-1} \{a^n \times (R_{1,k} + R_{2,k}) + b^n \times R_{3,k}\} \qquad (5)$$

$$a = e^{\frac{1}{t}} \qquad (6)$$

$$b = e^{-\frac{1}{t}} \qquad (7)$$

In (6) and (7), coefficients a, b are proposed in this study. a is a correction for a premium for negative rewards and b is a correction for a discount for positive rewards. These coefficients take into account the potential increase in discrepancies between the SoC, PV output predictions, and price predictions when the number of actions t increases. This consideration is important because larger discrepancies can occur between the predicted and actual values, leading to higher errors. The total sum of rewards obtained in a single learning process is given by (8), where N is the total number of steps in one training session and $days$ is the number of training days, as shown in (9):

$$R_{total} = \sum_{k=1}^{N} R_k \qquad (8)$$

$$N = days \times \frac{48}{t} \qquad (9)$$

III. Case Study

In the Case Study, a simulation is used to develop a charge/discharge plan for the actual study of battery control at AIST in the future. PV output and energy prices are predicted based on the acquisition of weather forecast data, and the results of the charge/discharge plan are output based on the predicted results.

The training data used consists of the observation data from the Japan Meteorological Agency's (JMA) Surface and Areological Observatory located in Tsukuba City, adjacent to

AIST. The data covers the period from April 1, 2022, to March 31, 2023. The input data for the forecast of future weather was obtained from the Grid Point Value (GPV) data produced by JMA and released as open data by the Research Institute for Sustainable Humanosphere (RISH) of Kyoto University. The GPV data used takes into consideration that the spot market closes at 10 a.m. (JST) of the preceding day of operation and that it takes a few hours from observation to publication as shown in Fig. 5. Therefore, the forecast for the most recent 78 hours observed at 9 p.m. (JST) is utilized. As there was no measured PV output data available from the experimental facility, simulated data was generated using the numerical values from the training dataset for this study. The sources of data acquisition are listed in Table. 2, and various parameters are presented in Table. 3.

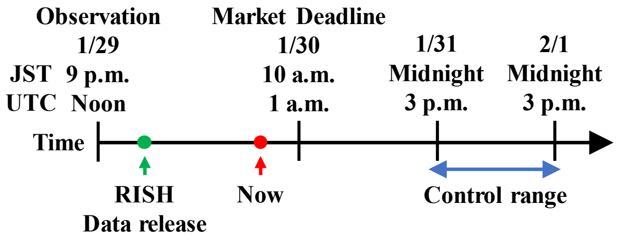

Fig. 5 Time sequence

TABLE. 2 DATA SOURCES

Parameter	Source
Temperature, Precipitation, Solar radiation, pressure, humidity	Areological Observatory, JMA, Tsukuba City
GPV	RISH
PV generation	self-made
Energy trading price	Japan Electric Power exchange
Imbalance fee	Imbalance prices Calculation Service

As an example of the simulation results, focusing on a sunny day when PV output is stable and predictable, compared to rainy and cloudy days. The results for January 31, 2023, are shown in Fig. 6, which displays the predicted PV output and the actual measurements. Fig. 7 illustrates the predicted energy prices. Additionally, the charge/discharge plan for the battery throughout the day will be presented. The results for the most feasible case of PV output are also shown in Fig. 8 for the charge/discharge plan of the battery in one day.

TABLE. 3. PARAMETERS

Times of study	40000
Days of study	365
Days of test	1
North latitude	140.134985
East longitude	36.064898

Fig. 6 represents the PV output [kW] on the vertical axis and time [h] on the horizontal axis. The solid lines represent the "lower" and "upper" bounds of the 95% prediction interval, while the dotted line represents the observed values. It can be observed that all actual PV outputs fall within the range of the 95% prediction interval. While it is a good result that everything falls within the prediction interval, ideally, the width of the prediction interval should be narrower and "observed" should not fall outside of the prediction interval.

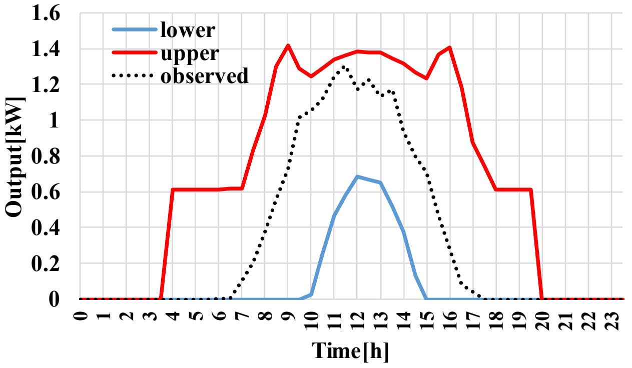

Fig. 6 Forecasted PV generation for a day

Fig. 7 represents the price per hour [JPY/kWh] on the vertical axis and time [h] on the horizontal axis. The solid line represents the predicted values, the dotted line represents the observed values, the red line represents the energy trading price during selling, and the blue line represents the imbalance fee. Although accuracy was not the main focus in this case, the RMSE values for the energy trading price and the imbalance fee are 1.565 and 6.390, respectively. The weight of the reward and penalty will change depending on the error, which is expected to affect the timing of the operation in the charge and/or discharge scheduling and will be discussed in the future.

In Fig. 8, the vertical axis represents the charge/discharge output of the battery [kW], the energy transaction price per 30 minutes [JPY/kWh/30min], and SoC [%], and the horizontal axis represents time. The charge/discharge plan is formulated with profitability in mind, with maximum charging before 3 p.m., when energy prices are low and PV is generating energy, and discharging all at once after 5 p.m., when energy prices are relatively high. The final net profit for the day was 156.1 [JPY].

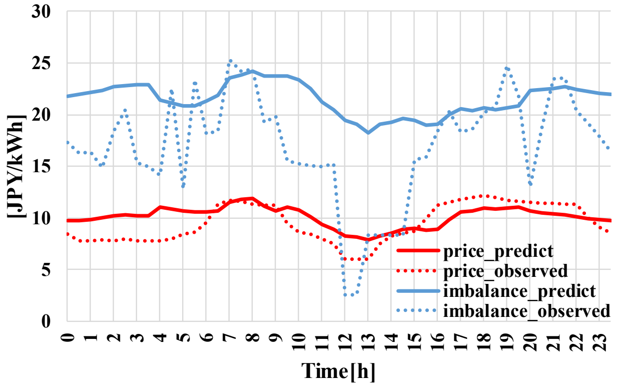

Fig. 7 Price and imbalance fee

On the other hand, the net profit was 159.1 [JPY] when operating with the same model, assuming that all the PV output and energy price forecasts were on target, which is the ideal outcome we are aiming for in the future. Although the values are close as revenue values, there is room for improvement since they are coincidental results based on forecasting errors. Compare how much the charge/discharge plan using the lower and upper PV output forecasts deviates from the ideal forecast with no error.

The charge and discharge scheduling for the battery throughout the day will be presented in Fig. 9. In Fig. 9, the left y-axis represents the actual energy price [JPY/kWh], the right y-axis represents SoC [%], and the x-axis represents hours in a day. The gray area is probabilistic interval. The gray area is generated based on the probabilistic PV and energy price forecasting with 95 [%] confidence. The charge and/or discharge scheduling for a battery is formulated with profitability in mind, with maximum charging before 3 p.m., when energy price is lower, and discharging all at once after 5 p.m., when energy price is relatively higher than other periods. The final net profit for the day was 156.1 [JPY]. On the other hand, the net profit was 159.1 [JPY] when operating with the same model, assuming that all the PV generation and forecasted energy price were on target, which is the ideal outcome. Although the values are close as revenue values, there is room for improvement since they are coincidental results based on forecasting errors. Compare how much the charge and/or discharge schedule using probabilistic PV generation forecast deviates from the ideal forecast with no error.

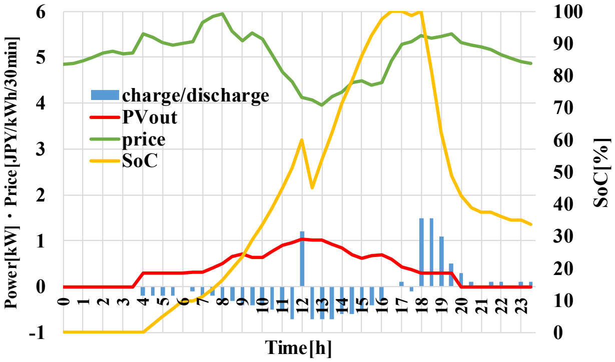

Fig. 8 charge and/or discharge scheduling

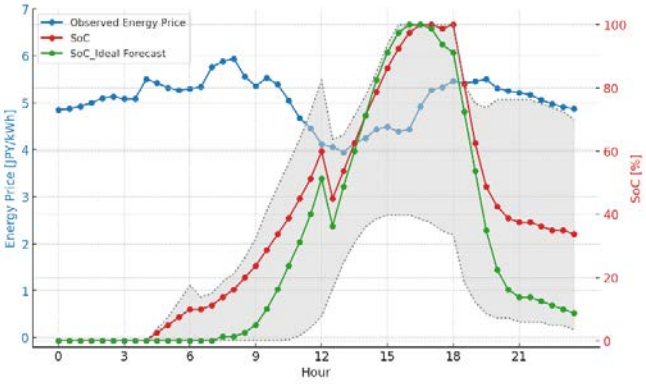

Fig. 9 SoC transition with 95% probability: observed PV generation vs ideal forecast for PV generation

IV. CONCLUSION

In this study, we proposed deep RL-based battery scheduling for a prosumer. The proposed model is designed to maximize a prosumers profit via day-ahead energy trade market. By the virtue of probabilistic forecasting of PV generation and energy price in the market, the case study shows the proposed method evaluates the risk of imbalances and maximizes the profit of the prosumer. In the future, we aim to operate the actual device, and verify whether the proposed RL model works properly.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number JP23K13306.

REFERENCES

[1] Inc. Tokyo Electric Power Company Holdings, “Open Platform Aggregation Business Demonstration Project Tokyo Electric Power Consortium Subsidy for Demonstration Project for Construction of Virtual Power Plant Using Demand-Side Energy Resources,” Mar. 2020. Accessed: Jun. 06, 2023. [Online]. Available: https://sii.or.jp/vpp31/uploads/B_1_2_tepco.pdf

[2] Z. Wu et al., “Real-Time Scheduling Method Based on Deep Reinforcement Learning for a Hybrid Wind-Solar- Energy Generation System,” 2021.

[3] M. Khan, J. Seo, and D. Kim, “Real-Time Scheduling of Operational Time for Smart Home Appliances Based on Reinforcement Learning,” IEEE Access, vol. 8, pp. 116520–116534, 2020, doi: 10.1109/ACCESS.2020.3004151.

[4] H. M. Abdullah, A. Gastli, and L. Ben-Brahim, “Reinforcement Learning Based EV Charging Management Systems-A Review,” IEEE Access, vol. 9. Institute of Electrical and Electronics Engineers Inc., pp. 41506–41531, 2021. doi: 10.1109/ACCESS.2021.3064354.

[5] H. van Hasselt, A. Guez, and D. Silver, “Deep Reinforcement Learning with Double Q-learning,” Sep. 2015, [Online]. Available: http://arxiv.org/abs/1509.06461

[6] B. Huang and J. Wang, “Deep-Reinforcement-Learning-Based Capacity Scheduling for PV-Battery System,” IEEE Trans Smart Grid, vol. 12, no. 3, pp. 2272–2283, May 2021, doi: 10.1109/TSG.2020.3047890.

[7] H. Yamamoto, J. Kondoh, and D. Kodaira, “Assessing the Impact of Features on Probabilistic Modeling of Photovoltaic Power Generation,” Energies (Basel), vol. 15, no. 15, Aug. 2022, doi: 10.3390/en15155337.