Journal of Electrical Engineering & Technology (2022) 17:727–735
https://doi.org/10.1007/s42835-021-00881-8

ORIGINAL ARTICLE

An Optimal Charging and Discharging Scheduling Algorithm of Energy Storage System to Save Electricity Pricing Using Reinforcement Learning in Urban Railway System

Hosung Jung[1]

Received: 6 May 2021 / Revised: 8 August 2021 / Accepted: 18 August 2021 / Published online: 3 September 2021

Abstract
This paper proposes the optimal charging and discharging scheduling algorithm of energy storage systems based on reinforcement learning to save electricity pricing of an urban railway system in Korea. Optimization is done through reinforcement learning of charging and discharging schedule of energy storage systems according to the unit of electricity pricing rates as well as a reduction of peak power demand to save electricity pricing. To do this, modeling of urban railway systems including energy storage systems, electricity pricing rates, and changes in rates according to operations of energy storage systems are carried out. Reinforcement learning for an agent is also done to reduce peak power demand through DQN algorithm. Operation data of actual lines of urban railways operating with energy storage systems are utilized for learning. For this reinforcement learning, about 399(45.3%) incorrect data are removed and 481(54.7%) normal data are extracted. Through the reinforcement learning, maximum peak power demand is reduced by a targeted amount, 100 kW, from 2,982.4 kW to 2,882.4 kW. When the peak power demand is under 2,600 kW, charging at times when the power rate is cheaper and discharging at times when the power rate is more expensive are carried out, thus saving the total electricity pricing.

Keywords Electricity pricing · Energy storage system · Optimal charging and discharging scheduling algorithm · Reinforcement learning · Urban railway system

1 Introduction

Urban railway system consists of rolling stock load and station load. The greatest amount of power used is for operation of large capacity rolling stock load. Power used for rolling stock load depends on the train schedule. Peak power demand occurs during rush hours when there are more train services but the amount of power used is dramatically reduced during the dawn and midnight hours when there are fewer train services, which reduces the total load for the system.

The industrial pricing rate is applied for the electricity pricing of the urban railway system in Korea. This is composed of usage rate calculated from hourly usage and basic rate calculated from peak power demand in 15-min periods. For the basic rate, monthly peak power demand from the current month and the past 12 months are compared and calculated by setting the biggest value as the standard. Therefore, the peak power demand impacts the basic rate for 12 months henceforth, even if it happened only once [1–3].

Because reducing the amount of power used through rolling stock load control has some restrictions, peak power demand is being controlled through limiting the power supply to a load system that uses a relatively larger amount of power among the station load system. However, this method also has some restrictions in reducing the peak power demand because of environmental issues. Therefore, the urban railway system manages peak power demand through an energy saving device to reduce the peak power demand. Charging the energy storage system (Energy Storage System, ESS) with power during the light duty times and discharging it during the peak power times is being used nowadays [5–7]. The energy storage system allows charging the power and discharging it during the desired times. Therefore, it is being used for profit-taking in time-specific rates, volatility

✉ Hosung Jung
hsjung@krri.re.kr

[1] Smart Electrical & Signaling Division, Korea Railroad Research Institute (KRRI), 176, Cheoldo Bangmulgwan-ro, Uiwang-si 16105, Gyeonggi-Do, Korea

control of renewable power generation, and reducing peak power demand [8–10].

Previous studies about optimal scheduling methodology that establish hourly charging and discharging schedules for an energy storage system using the 24-h demand forecast indicated ways to reduce the peak power demand in particular. This method uses linear programming or integer programming to offer an optimal schedule that can maintain low electricity pricing while satisfying all the constraints. However, any error in the forecast will be critical in reducing the peak power demand. A study is undertaken to control the energy storage system through an algorithm that repeats short-term forecast and rescheduling in virtual power plants. However, this method can prepare for short-term volatility but has some restrictions in preparing for long-term changes in rates [11–14].

Therefore, this paper focuses on reinforcement learning approach among artificial intelligence approaches applied to various industries for optimal scheduling of ESS to reduce peak power demand. Reinforcement learning is a type of machine learning and the agent learns behavior that maximizes reward through repeating trial and error. It can be applied to an environment where mathematical modeling is difficult to build [15–17].

This paper proposes an optimal charging and discharging scheduling algorithm of ESS to reduce peak power demand in urban railway systems using the DQN (Deep Q-Network) method. Modeling for the urban railway system including ESS is made and a reinforcement learning environment is formed to reduce the peak power demand. Then the agent learned in the environment and the effect is analyzed. Power data from real operation lines with ESS are used for reinforcement learning data. Abnormal data are removed and preprocessing is performed to make the data suitable for learning.

2 Optimization Algorithm for Electricity Pricing Using DQN Algorithm

2.1 Urban Railway System Modeling for Reinforcement Learning

Electricity pricing for urban railway systems is calculated according to pricing rates from the power supplier, which depend on the amount of power used by rolling stock loads and station loads as shown in Fig. 1. Generally, an industrial pricing rate is applied to urban railways in Korea. The industrial pricing rate is composed of a basic rate calculated from peak power demand and usage rate calculated from hourly usage. A power supplier supplies power to a substation where the amount of power used by rolling stocks in the section and amount of power used by stations that receive power from the substation are added to calculate the amount of power used by the urban railway system. Recently, an energy storage system is applied to use the power at the appropriate time for the most effective usage of regenerative power and renewable energy. Energy storage systems not only reduce peak power demand but also reduce electricity pricing by charging power at hours when the rate is cheaper and discharging at more expensive hours.

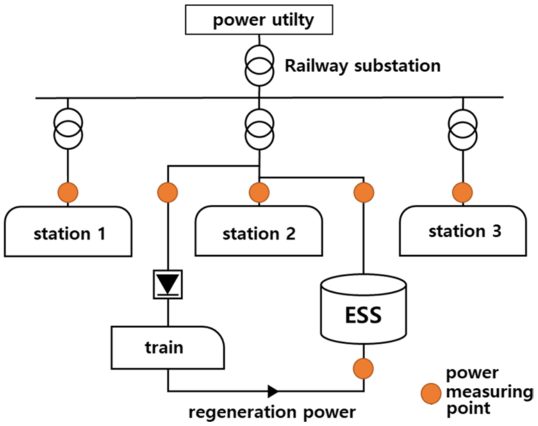


Fig. 1 Urban railway system model for reinforcement learning

This paper uses reinforcement learning to reduce electricity pricing of urban railway systems by setting optimal charging and discharging schedule with an energy storage system. For this we define urban railway system's environment state, action, and reward of an urban railway system operating energy storage system as shown in Fig. 1 to apply reinforcement learning.

First of all, the environment state of the urban railway system shown in Fig. 1 can be defined as in Eq. 1.

$$s_t = < P_t^{Load}, SoC_t, t, C_t^E > \tag{1}$$

s_t stands for the state at time t, P_t^{Load} stands for total demand at time t, and total demand is calculated from rolling stocks load, stations load, and charge/discharge of energy storage system (Total demand = Rolling stocks demand + Stations demand + Charge/Discharge of energy storage system). SoC_t is energy charging ratio to the maximum capacity of the energy storage system at time t. t stands for time. Power consumption and the charging/discharging pattern of energy storage systems are shown to change according to the operation time. Each value shows a certain pattern according to the operating time. Lastly, C_t^E refers to the unit cost of power. The power rates may differ depending on detailed rates but generally, it is classified into three steps, light load, medium load, and heavy load.

The learning agent carries out charge or discharge and maintains the present stage through actions. It uses discrete charging/discharging amounts to reduce the complexity of a problem. Real operation lines data is used in this paper. The maximum charging/discharging amount of the energy storage system in these lines is 200 kW. Therefore, Eq. (2) is defined to show a set of actions that an agent can take.

$$a_t = < -200, -100, 0, 100, 200 \ > \tag{2}$$

Fundamentally, reward has a correlation with electricity pricing rate of each time. To effectively represent the actions of energy storage systems, changes in pricing rates according to charge/discharge of the energy storage system as shown in Eq. (3) is used instead of electricity pricing rates.

$$R\left(S_t, A_t\right) = C_t A_t / 4 \tag{3}$$

$R\left(S_t, A_t\right)$ refers to reward when action A_t is chosen for state S_t at time t. C_t stands for the rate at each time which refers to the actual electricity unit price. However, because calculation of peak power demand is done in 15-min periods in urban railway systems, the reward is divided by 4. In addition, if the energy storage system repeats charge/discharge in a short period of time when the same rate is applied, actual electricity pricing can increase due to the losses and it can shorten the life of the equipment due to frequent charging/discharging. Therefore, to prevent frequent charging/discharging, a penalty (Φ)is given when the multiples of A_t and A_{t-1} are less than 0 as shown in Eq. (4) which means when an energy storage system changes the operation state from charging to discharging or discharging to charging. In this way, the agent's action can learn in a way for reward to be increased.

$$\text{If)}A_t \bullet A_{t-1} < 0 \tag{4}$$

$$R\left(S_t, A_t\right) + = \Phi$$

The state transition from time t to time t + 1 is carried out as follows. P_{t+1}^{Load} Can be provided through external predicted value. SoC_{t+1} uses the calculated value from actions within the capacity range (Cap) of the energy storage system at its existing SoC_t state as shown in Eq. (5). Time $t + 1$ uses the value next to the current time t value in a day cycle. $t + 1$ is provided after calculating the rate level at that time C_{t+1}^E.

$$SoC_{t+1} = SoC_t + \frac{A_t}{Cap} \tag{5}$$

Modeling of reward and state transition way is done to carry out the modeling of the urban railway system's environment for reinforcement learning.

2.2 DQN(Deep Q-Network) Algorithm

DQN algorithm is applied in this paper to learn a learning agent in the urban railway system's environment. DQN approximates function Q to DNN (Deep Neural-Network). There are a lot of state sets so it can solve difficult problems that cannot be solved with existing algorithms. It uses CNN (Convolution Neural-Network) to carry out reinforcement learning with image, voice, and various data. Also, DQN uses replay memory to speed up the learning process by removing the correlation between samples when carrying out reinforcement learning. It also uses a target network to increase the efficiency of learning. Figure 2 shows the learning process of the DQN algorithm. Function Q approximates an artificial neural network to carry out learning of function Q. Experience replay is used to remove the temporal correlation between samples used for learning. Updates using numerous samples are carried out to enable stable learning. In the DQN algorithm, the solution regarding states that have successive values is possible since function Q is approximated to an artificial neural network [18–20].

DQN reinforcement learning algorithm is based on the Markov decision process. Therefore, it is important to model the learning environment, actions that an agent can carry out, and the following rewards appropriately. Also, preprocessing of learning data is necessary to test and learn from data collected from real operation lines.

3 Urban Railway Power Data Preprocessing

3.1 Outlier Elimination Method

Outliers in acquired power consumption data as shown in Fig. 3 can occur from various causes such as sensor error and communication error in real operating lines. These Outliers can occur throughout the day or at a certain time.

In this paper, standard distribution of power data from the past 2 years has been proposed and for those that deviate from normal distribution over time in a time series, the

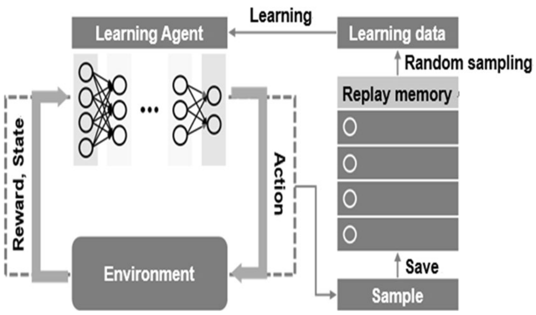


Fig. 2 DQN algorithm's learning process

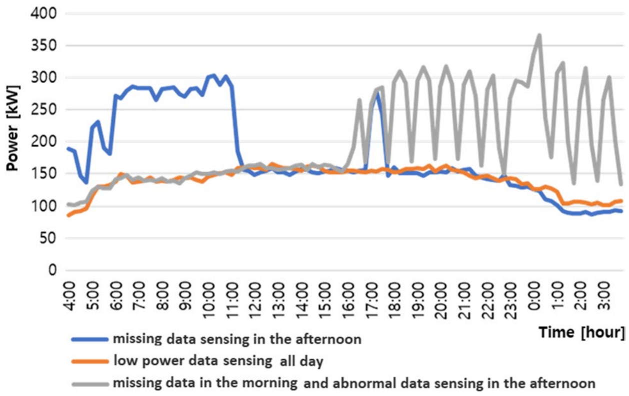


Fig. 3 Data samples including abnormal data

power consumption data are judged to be outlier data and removed.

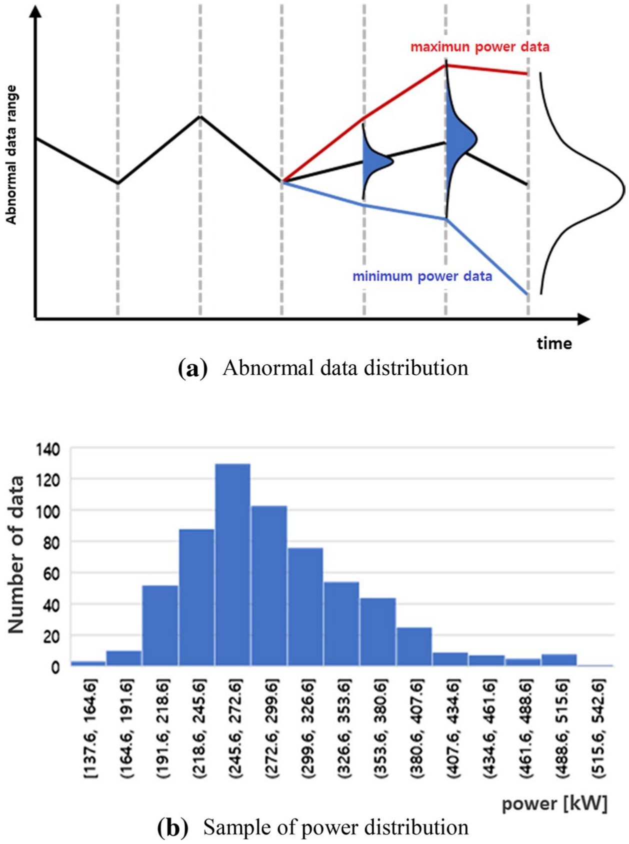
It is to carry out the reinforcement learning on a day-to-day basis. Figure 4 presents an outlier data detection method out of time series data. However, actually measured power consumption data by time step do not follow a normal distribution. Therefore, daily demand is patterned through normalization to judge by focusing on patterns rather than demand values by showing data in values between 0 and 1 through data normalization. Equation (6) used min–max normalization method and it is a process of finding the normalized values for each hour [21].

$$\overline{P_t^{Load}} = \frac{P_t^{Load} - P_{min}^{Laod}}{P_{max}^{Load} - P_{min}^{Load}} \tag{6}$$

$\overline{P_t^{Load}}$ is normalized power value from time t, P_{max}^{Load} and P_{min}^{Load} represents maximum and minimum power values for the day. Figure 5 shows standard patterns through normalized power data after eliminating outliers.

Preprocessing of learning data is performed by eliminating data for the days when data that exceeds the standard value occur as shown in Eq. (7) once. The standard power consumption pattern is established as shown in Fig. 5. However, the power consumption pattern differs according to seasons. Therefore, the standard pattern is continuously updated as shown in Eq. (8) using the actually measured data.

$$\left| \overline{P_t^{Load}} - \widehat{P_t^{Load}} \right| > \varepsilon \tag{7}$$



(a) Abnormal data distribution

(b) Sample of power distribution

Fig. 4 Detecting outliers in time series power data

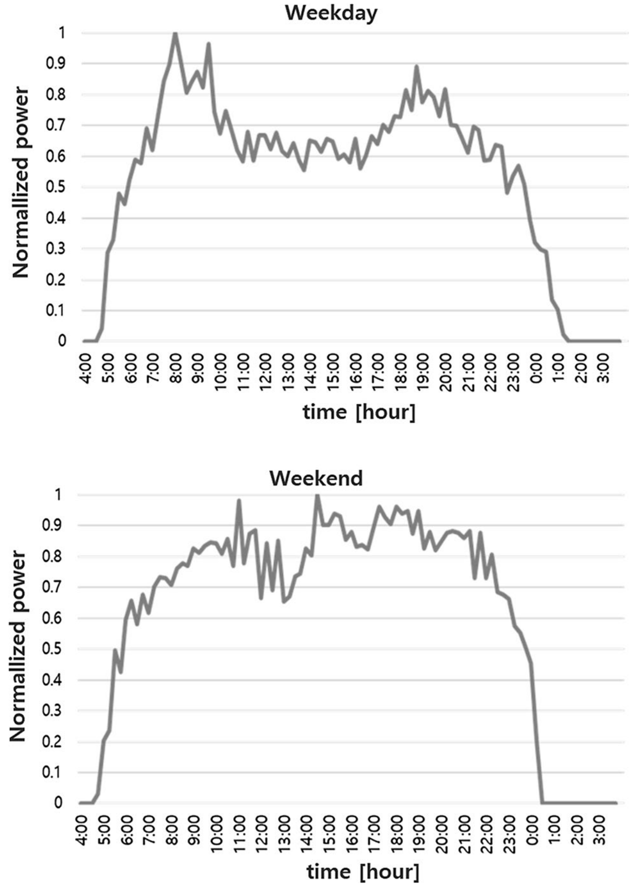


Fig. 5 Standard power data pattern (weekday, weekend)

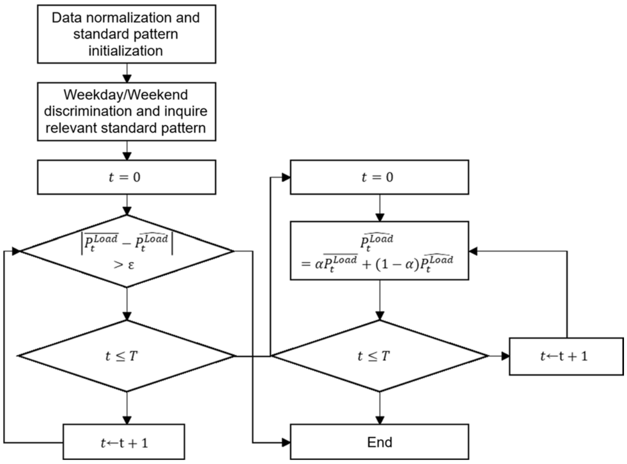

Data normalization and standard pattern initialization
Weekday/Weekend discrimination and inquire relevant standard pattern
$t = 0$
$|\overline{P_t^{Load}} - \widehat{P_t^{Load}}| > \varepsilon$
$t \le T$
$t \leftarrow t + 1$
$t = 0$
$\widehat{P_t^{Load}} = \alpha\overline{P_t^{Load}} + (1-\alpha)\widehat{P_t^{Load}}$
$t \le T$
$t \leftarrow t + 1$
End

Fig. 6 Flow chart of outlier elimination algorithm

Table 1 Outlier elimination ratio

	Total data	Removed data	Remained data	Removed ratio
Number of data	880	399	481	45.3%

$$\widehat{P_t^{Load}} = \alpha\overline{P_t^{Load}} + (1-\alpha)\widehat{P_t^{Load}} \tag{8}$$

α is a factor that takes a value between 0 and 1, and the larger the value, the closer the standard pattern is α to the recent load pattern.

Figure 6 shows an overall flow chart showing the preprocessing process of urban railway power data for reinforcement learning.

3.2 Result of Outlier Elimination

Table 1 shows the number of data entries after eliminating outliers from the number of daily power consumption data from operation lines and percentage of final learning data available for learning applying the outlier elimination algorithm. About 45% of all data is eliminated. Figure 7 shows the remaining daily learning data samples according to the power consumption pattern of summer (August) and winter seasons (December). It shows that the amount of power consumption increases in daytime during summer and in the evening/nighttime during winter.

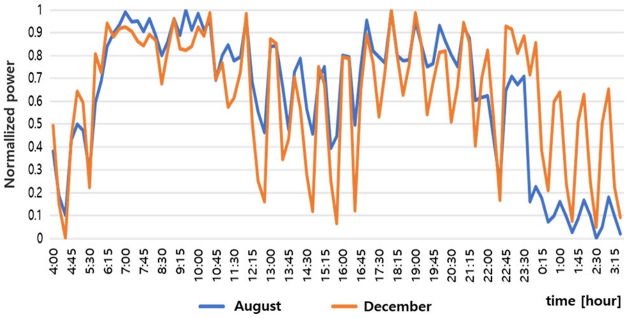


Fig. 7 Remained data samples (summer, winter)

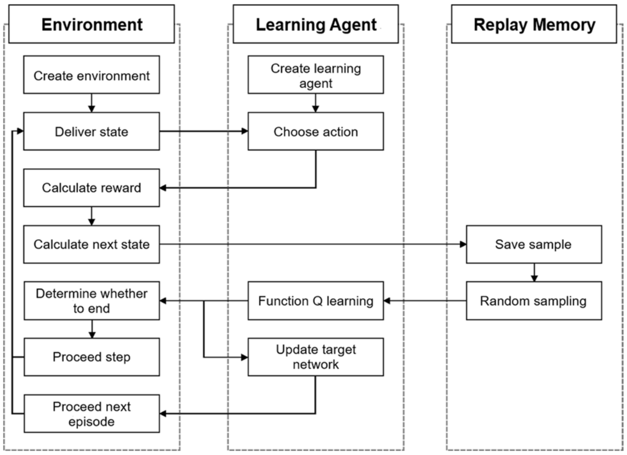


Fig. 8 Learning process of DQN algorithm

4 Electricity Pricing Saving Performance Evaluation

4.1 Learning Process

Figure 8 shows the learning process of the reinforcement learning method that is applied for optimization of urban railway system's electricity pricing. It consists of environment, learning agent, and replay memory.

First of all, the reinforcement learning method create a learning agent expressed as DNN representing function Q. The learning agent should include a neural network as the subject of learning and parameters required for learning. The learning agent will take an action on the state of the urban railway system, calculate the next state according to the reward, save the data in replay memory, randomly extract samples from stored samples to learn function Q, and judge whether to end the learning process or not. If learning does not end, the target network is updated, and learning is performed repeatedly by proceeding to the next step for the new episodes and having the agent perform actions according to the state.

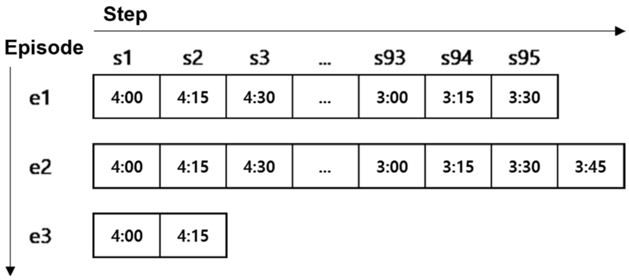


Fig. 9 Definitions of step and episode

Steps and episodes are defined as shown in Fig. 9. Step refers to 15-min periods that decide peak power demand. It is set for the step to start at 4:00 AM. Episode refers to a set of consecutive steps on a daily basis. Episode consists of steps until the episode termination condition is met. Therefore, the step length is different for each episode. In this paper, if there is a time period that exceeds the SoC range during the operation of energy storage system, control is performed until that time and the episode is terminated.

For reinforcement learning, filtered data are used after outlier elimination. Two-thirds of total data is randomly sampled to be used for agent training and the rest of the data are used for verification. 323 days of data are used for training and 158 days of data are used for verification.

Among the filtered data used for reinforcement learning, daily maximum peak power demand recorded 2,982.8 kW and the maximum discharge of energy storage system is 200 kW. Therefore, it would be appropriate to target 2,800 kW. However, if too large a P^{Peak} value is set, the number of days carrying out discharge will be insufficient and the learning may not be carried out properly. Hence, the target is set to 2,600 kW so that learning takes place on most measurement days. The penalty is set to 50,000 won(₩) when it went out of SoC range constraint. This value is set to a value larger than the value obtained by multiplying the maximum discharge amount of the energy storage device, 200 kW, and the highest rate plan unit price of 189.7 won(₩)/kW, resulting in the penalty being set higher than the reward due to discharge. The penalty for exceeding peak power demand is set to 1,000 won(₩) so that charging and discharging of the energy storage system can be performed efficiently. Also, the penalty for charging near the peak power demand is set to 5,000 won(₩) to prevent the energy storage system from charging near the peak power demand.

In this way, the score is calculated according to the learning agent's learning of episode. These scores have a different demand value for each episode. Therefore, the learning agent's level of learning is judged according to the moving average of scores instead of individual scores. In this paper, it is set to be terminated when the moving average of scores exceeds 20,000 won(₩). This takes into an account the case of discharging 400 kW which is 89% of total capacity in seasons with little difference in rates (spring, fall).

4.2 Learning and Verification Result

Figure 10 shows the learning progress of the reinforcement learning model. It represents the moving average scores according to episodes. In the early stage of learning, the energy storage system is randomly activated in order to accumulate enough data in the replay memory, so the episode ends while exceeding the SoC constraints within a short period of time and the number of steps in an episode is significantly less than in the beginning part of learning. The score increased dramatically after learning started and we are able to see that it converged toward the target value at the 313th episode.

Figure 11 shows changes in peak power demand before and after reinforcement learning. The maximum peak power demand went from 2,982.4 kW to 2,882.3 kW, which satisfied the targeted value of reducing 100 kW. In addition, it is trained to reduce the overall electricity pricing by discharging the power stored in the energy storage system during the peak times to reduce the peak power demand on the days when peak power demand is over 2,600 kW. While, on the days when the peak power demand is under 2,600 kW, the system is trained to charge in the morning time, when the power unit price is cheaper, and discharge at heavy load hours rather than to reduce the peak power demand. Thus, the overall daily peak power demand tends to increase in days when the peak power demand is lower. This occurs

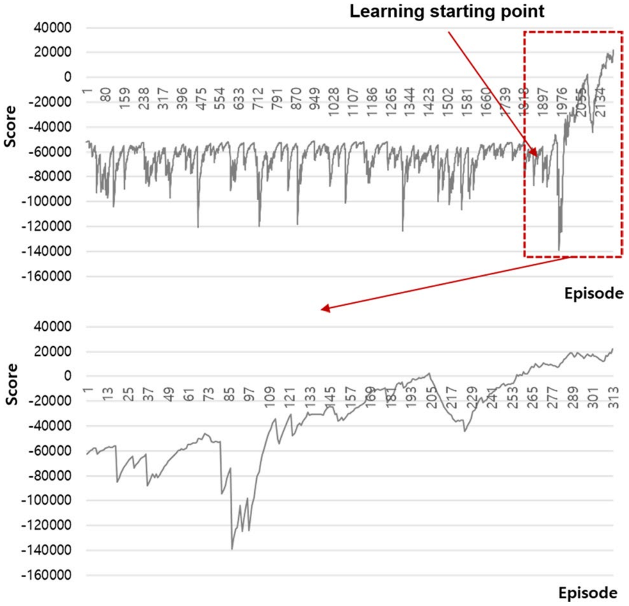


Fig. 10 Moving average value of scores according to progress of episodes during learning

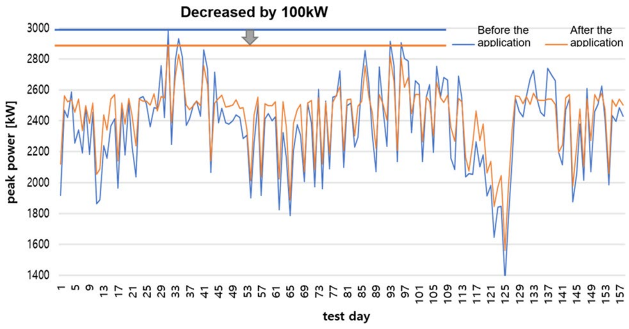


Fig. 11 Changes in peak power demand before and after the application of reinforcement learning (Unit: kW)

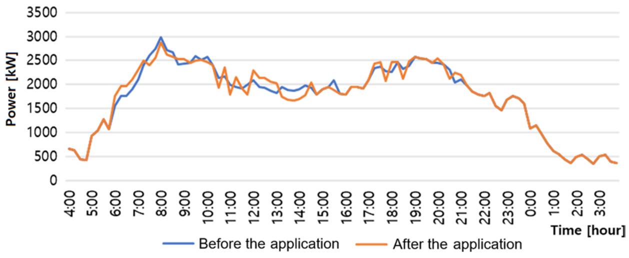


Fig. 12 Power consumption before and after reinforcement learning on days when peak power demand occurred (Unit: kW)

because the energy storage system is set to start charging when load exceeds 1,500 kW. On the days when the power consumption increases gradually due to low peak power demand, it takes about 2 h to fill up SoC. Therefore, it starts charging at around 7:00 AM, which is normally during peak power demand hours and hence, increases the peak power demand. However, an increase in peak power demand does not affect overall electricity pricing when the peak power demand is low. Hence, the reinforcement learning model shows the learning result that reduces overall electricity pricing.

Figure 12 shows power consumption graph on the days when peak power demand occurred during the entire period. When compared with the existing charging and discharging scenario to minimize the maximum peak power demand, it is confirmed that peak power demand is reduced by starting charging at about 6:00 AM and starting discharging at about 7:30 AM when peak power demand occurred.

Figure 13 shows unit price per time step and ESS SoC on the days when peak power demand occurred. Charging started at about 6:00 AM and discharging started at about 7:30 AM during the peak power demand hours to reduce peak power demand. The figure also shows that small charging and discharging is repeated during the medium load and heavy load hours to reduce electricity pricing according to electricity pricing rates. And ESS charges and discharges in ways to save overall electricity pricing by charging during the medium load hours in lunch hours and discharging during the heavy load hours in the afternoon.

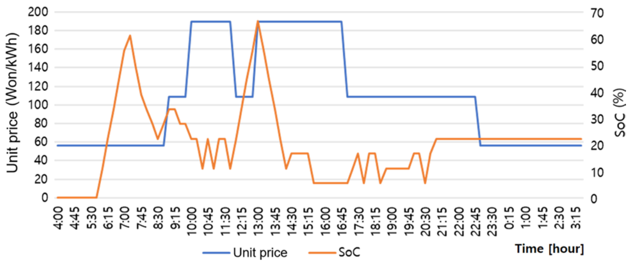


Fig. 13 Hourly unit prices and SoC changes on the days when peak power demand occurred

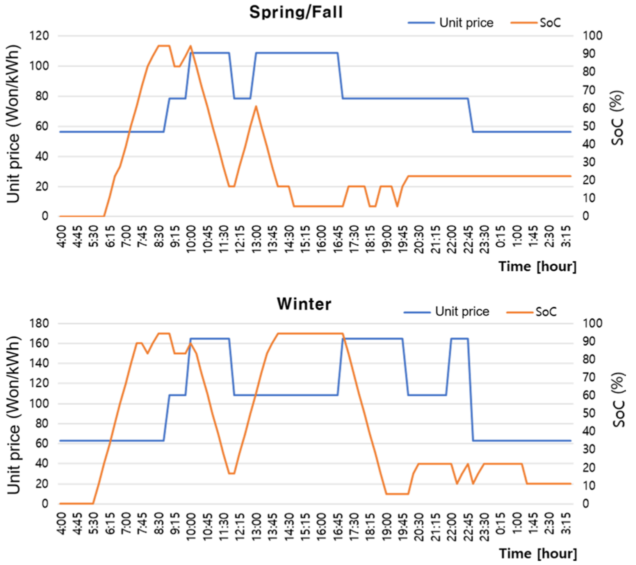


Fig. 14 Changes in SoC and hourly unit prices in spring/fall and winter

Figure 14 shows changes in SoC when peak power demand did not occur. Spring, fall, and winter seasons have relatively lower peak power demand than summer (peal power demand is 2,422.8 and 2,425.2 kW respectively). Therefore, charging and discharging of the energy storage system is carried out in the direction of lowering the electricity pricing rather than reducing peak power demand. It is confirmed that charging and discharging is generally carried out in the direction of lowering the electricity pricing by charging during light load hours or medium load hours and discharging during heavy load hours.

5 Conclusion

This paper proposes the optimal charging and discharging scheduling algorithm of energy storage system based on reinforcement learning to reduce electricity pricing of urban railway systems in Korea. To do this, modeling of an urban railway system including its energy storage system, electricity pricing rates applied, and changes in rates according to actions of the energy storage system are carried out. Reinforcement learning agent is also performed to reduce peak power demand through DQN algorithm. An energy storage device is installed for the purpose of reinforcement learning and the actual operating lines data of the urban railway is used. For more effective learning, two years data measured from real operation lines are preprocessed to present a standard power consumption pattern, and outliers found out of the range of the standard model is eliminated through time series comparison.

For the measurement samples, approximately 399 (45.3%) outliers are eliminated, and 481 (54.7%) normal data are extracted and used for reinforcement learning.

Through the reinforcement learning using data from actual line operation, the maximum peak power demand within the entire period decreased from 2,982.4 kW to 2,882.4 kW, confirming that the targeted reduction of 100 kW is successfully accomplished. In addition, it is confirmed that the energy storage system operated in the direction of reducing the overall electricity pricing by discharging the power stored in the energy storage system during the peak times to reduce the peak power demand on the days when peak power demand is over 2,600 kW and by charging when the electricity pricing rate is cheaper and discharging when the electricity pricing rate is expensive rather than reducing the peak power demand on days when the peak power is less than 2,600 kW.

The reinforcement learning-based charging/discharging scheduling algorithm of energy storage system helps to derive an optimized charging/discharging scenario in a complex system such as an urban railway, and once the learning is completed, immediate results can be derived even if the environment changes. It is, therefore, expected to be suitable for deriving charging/discharging scenarios for the energy storage system in the actual operation lines. However, the proposed reinforcement learning model is still under development. Thus, future studies with the latest reinforcement learning models based on data from various operation lines required.

Acknowledgements This research was supported by Korea Railroad Research Institute, Republic of Korea.

References

1. Zhu F, Yang Z, Xia H, Lin F (2018) Hierarchical control and full-range dynamic performance optimization of the supercapacitor energy storage system in urban railway. IEEE Trans Ind Electron 65(8):6646–6656. https://doi.org/10.1109/TIE.2017.2772174
2. Ratniyomchai T, Hillmansen S, Tricoli P (2014) Recent developments and applications of energy storage devices in electrified railways. IET Electr Syst Transp 4(1):9–20. https://doi.org/10.1049/iet-est.2013.0031
3. Lee H, Lee H, Lee C, Jang G, Kim G (2010) Energy storage application strategy on DC electric railroad system using a novel railroad analysis algorithm. J Electr Eng Technol 5(2):228–238. https://doi.org/10.5370/JEET.2010.5.2.228,[Online]
4. Ratniyomchai T, Hillmansen S, Tricoli P (2013) Recent developments and applications of energy storage devices in electrified railways. IET Electr Syst Transp 4(1):9–20. https://doi.org/10.1049/iet-est.2013.0031,[Online]
5. P. Radcliffe, J. S. Wallace and L. H. Shu. (2010). Stationary applications of energy storage technologies for transit systems. In: 2010 IEEE Electrical Power & Energy Conference, Halifax, NS. pp. 1–7. [Online]. DOI: https://doi.org/10.1109/EPEC.2010.5697222
6. Lee H, Kim J, Lee C et al (2020) novel cooperative controller design of heterogeneous energy storages for economic applications in electric railway systems. J Electr Eng Technol 15:979–987. https://doi.org/10.1007/s42835-019-00341-4,[Online]
7. Xia H, Chen H, Yang Z, Lin F, Wang B (2015) Optimal energy management location and size for stationary energy storage system in a metro line based on genetic algorithm. Energies 8(10):11618–11640. https://doi.org/10.3390/en81011618
8. Mian Qaisar S (2020) A proficient Li-ion battery state of charge estimation based on event-driven processing. J Electr Eng Technol 15:1871–1877. https://doi.org/10.1007/s42835-020-00458-x
9. Ruiz-Corts M et al (2019) Optimal charge/discharge scheduling of batteries in microgrids of prosumers. IEEE Trans Energy Convers 34(1):468–477. https://doi.org/10.1109/TEC.2018.2878351
10. V. I. Herrera, H. Gaztaaga, A. Milo, A. Saez-de-Ibarra, I. Etxeberria-Otadui and T. Nieva. (2015). Optimal energy management of a battery-supercapacitor based light rail vehicle using genetic algorithms. In: Proc. IEEE Energy Convers. Congr. Expo., pp. 1359–1366. [Online]. DOI: https://doi.org/10.1109/ECCE.2015.7309851
11. Vazquez S, Lukic SM, Galvan E, Franquelo LG, Carrasco JM (2010) Energy storage systems for transport and grid applications. IEEE Trans Ind Electron 57(12):3881–3895. https://doi.org/10.1109/TIE.2010.2076414
12. R. Barrero, X. Tackoen and J. V. Mierlo. (2008) Improving energy efficiency in public transport: stationary supercapacitor based energy storage systems for a metro network. In: Proceedings IEEE Vehicle Power Propulsion Conference, pp. 1–8. DOI: https://doi.org/10.1109/VPPC.2008.4677491
13. Ko R, Jo HC, Joo SK (2019) Energy storage system capacity sizing method for peak-demand reduction in urban railway system with photovoltaic generation. J Electr Eng Technol 14:1771–1775. https://doi.org/10.1007/s42835-019-00203-z
14. de la Torre S, Racero AJS, Aguado JA, Reyes M, Martínez O (2015) Optimal sizing of energy storage for regenerative braking in electric railway systems. IEEE Trans Power Syst 30(3):1871–1877. https://doi.org/10.1109/TPWRS.2014.2340911
15. Zhu F, Yang Z, Lin F, Xin Y (2020) Decentralized cooperative control of multiple energy storage systems in urban railway based on multiagent deep reinforcement learning. IEEE Trans Power Electr 35(9):9368–9379. https://doi.org/10.1109/TPEL.2020.2971637
16. Zhang D, Han X, Deng C (2018) Review on the research and practice of deep learning and reinforcement learning in smart grids. CSEE J Power Energy Syst 4(3):362–370. https://doi.org/10.17775/CSEEJPES.2018.00520
17. Nguyen TT, Nguyen ND, Nahavandi S (2020) Deep reinforcement learning for multi-agent systems: a review of challenges solutions and applications. IEEE Trans Cybern 50(9):3826–3839. https://doi.org/10.1109/TCYB.2020.2977374

18. Yang Z, Zhu F, Lin F (2021) Deep-reinforcement-learning-based energy management strategy for supercapacitor energy storage systems in urban rail transit. IEEE Trans Intell Trans Syst 22(2):1150–1160. https://doi.org/10.1109/TITS.2019.2963785
19. G. Palmer, K. Tuyls, D. Bloembergen and R. Savani. (2018). Lenient multi-agent deep reinforcement learning. In: Proc. 17th Int. Conf. Auton. Agents and MultiAgent Syst., pp. 443–451. [Online]
20. J. Foerster, N. Nardelli, G. Farquhar, P. H. S. Torr, P. Kohli and S. Whiteson. (2017). Stabilising experience replay for deep multi-agent reinforcement learning. In: Proc. 34th Int. Conf. Mach. Learning, pp. 1146–1155. [Online]
21. Ko R, Kong S, Joo SK (2015) Mixed integer programming (MIP)-based energy storage system scheduling method for reducing the electricity purchasing cost in an urban railroad system. Trans Korean Inst Electr Eng 64(7):1125–1129. https://doi.org/10.5370/KIEE.2015.64.7.1125,[Online]

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Hosung Jung received a B.S. and M.S. degree in Electrical Engineering from Sungkyunkwan University, Republic of Korea, in 1995 and 1998, respectively. He received a Ph.D. degree from the Electrical Electronic and Computer Engineering from Sungkyunkwan University in 2002. He is currently a chief Researcher with the Smart Electrical & Signaling Division, Korea Railroad Research Institute, Uiwang, South Korea.