Regular paper

Journal of the KIECS. pp. 93-104, vol. 19, no. 1, Feb. 29. 2024, t. 123, pISSN 1975-8170 | eISSN 2288-2189
http://dx.doi.org/10.13067/JKIECS.2024.19.1.93

SAC 강화 학습을 통한 스마트 그리드 효율성 향상: CityLearn 환경에서 재생 에너지 통합 및 최적 수요 반응

이자노브 알리벡 러스타모비치* · 성승제 · 임창균**

Enhancing Smart Grid Efficiency through SAC Reinforcement Learning: Renewable Energy Integration and Optimal Demand Response in the CityLearn Environment

Esanov Alibek Rustamovich* · Seung Je Seong · Chang-Gyoon Lim**

요 약

수요 반응은 전력망의 신뢰성을 높이고 비용을 최소화하기 위해 수요가 가장 많은 시간대에 고객이 소비 패턴을 조정하도록 유도한다. 재생 에너지원을 스마트 그리드에 통합하는 것은 간헐적이고 예측할 수 없는 특성으로 인해 상당한 도전 과제를 안고 있다. 강화 학습 기법과 결합된 수요 대응 전략은 이러한 문제를 해결하고 기존 방식에서는 이러한 종류의 복잡한 요구 사항을 충족하지 못하는 경우 그리드 운영을 최적화할 수 있는 접근 방식으로 부상하고 있다. 본 연구는 재생 에너지 통합을 위한 수요 반응에 강화 학습 알고리즘을 적용하는 방법을 찾아 적용하는데 중점을 둔다. 연구의 핵심 목표는 수요 측 유연성을 최적화하고 재생 에너지 활용도를 개선할 뿐 아니라 그리드 안정성을 강화하고자 한다. 연구 결과는 강화 학습을 기반으로 한 수요 반응 전략이 그리드 유연성을 향상시키고 재생 에너지 통합을 촉진하는 데 효과적이라것을 보여준다.

ABSTRACT

Demand response is a strategy that encourages customers to adjust their consumption patterns at times of peak demand with the aim to improve the reliability of the power grid and minimize expenses. The integration of renewable energy sources into smart grids poses significant challenges due to their intermittent and unpredictable nature. Demand response strategies, coupled with reinforcement learning techniques, have emerged as promising approaches to address these challenges and optimize grid operations where traditional methods fail to meet such kind of complex requirements. This research focuses on investigating the application of reinforcement learning algorithms in demand response for renewable energy integration. The objectives include optimizing demand-side flexibility, improving renewable energy utilization, and enhancing grid stability. The results emphasize the effectiveness of demand response strategies based on reinforcement learning in enhancing grid flexibility and facilitating the integration of renewable energy.

키워드

CityLearn, Soft Actor Critic Algorithm, Demand Response, Renewable Energy, Reinforcement Learning
시티런, 소프트 액터 크리틱 알고리즘, 수요 반응, 재생 에너지, 강화 학습

* 전남대학교 대학원 컴퓨터공학과(alli13@jnu.ac.kr)
전남대학교 대학원 컴퓨터공학과(sng_j@jnu.ac.kr)
** 교신저자 : 전남대학교 대학원 컴퓨터공학과

• 접 수 일 : 2023. 12. 29
• 수정완료일 : 2024. 01. 20
• 게재확정일 : 2024. 02. 17

• Received : Dec. 29, 2023, Revised : Jan. 20, 2024, Accepted : Feb. 17, 2024

• Corresponding Author : Chang-Gyoon Lim
Dept. of Computer Engineering, Chonnam National University,
Email : cglim@jnu.ac.kr

Ⅰ. Introduction

Transforming energy grids to become carbon-neutral demands radical alterations in how we consume energy, especially to cope with the fluctuating nature of wind and solar power generation. One approach to reduce this issue is demand response, a strategy that encourages users to modify their energy consumption from periods of low generation to times when energy generation is plentiful[1]. With the increasing adoption of renewable energy sources[2] like solar and wind, which are intermittent in nature, demand response can play a vital role in smoothing out the imbalances between energy supply and demand. Demand response is most often deployed through buildings and dealing with the uncertainties associated with RESs[3] to prevent instability and ensure resource availability, their integration into the current grid infrastructure must be done with care. The complexity of managing building controls increases because of the necessity to adaptively shift electrical loads in response to signals from the power grid[4]. Participating in demand response initiatives offers networks of buildings ways to improve control over energy loads and boost cost-effectiveness, all while easing the unpredictability of power inputs from renewable energy sources[5]. Several techniques have been proposed to optimize the energy consumption of grid-responsive buildings in order to match the power grid's demand. Among the various methods, model-based methods are the most researched. Specifically, model predictive control (MPC) has made considerable contributions to both energy management and demand response initiatives[6]. Another model-based approach involves modeling the demand response problem as a scheduling problem formulated as mixed-integer linear programming (MILP) problem, which requires knowledge of system dynamics for various appliances utilized in energy management systems[7]. Due to difficulties in dealing with time-varying system variables and the need to create distinct energy models for each building, the feasibility of using model-based methods for demand response diminishes as the scale of the problem expands.

As machine learning advances rapidly, reinforcement learning (RL) has shown considerable potential in addressing demand response as a series of decision-making steps in contrast to model-based approaches[8]. Unlike traditional optimization methods, RL doesn't necessitate pre-existing knowledge of system behavior and can be employed in a model-free way, simplifying its application in real-world scenarios. In recent years many RL methods have been proposed. Such as proximal policy optimization (PPO) algorithm[9], deep deterministic policy gradient (DDPG) off - policy algorithm[10], twin delayed DDPG (TD3) off-policy algorithm[11] and etc. RL has been successfully employed in developing several demand response programs[12 - 15]. In this research we use Soft Actor-Critic (SAC) algorithm, a state-of-the-art method in the domain of deep reinforcement learning, known for its sample efficiency and stability during training. SAC stands out due to its entropy regularization component, which encourages the exploration of various policy actions, leading to a more robust and comprehensive learning process. This intrinsic characteristic of SAC is particularly advantageous when dealing with the stochastic nature of renewable energy sources and the dynamic demands of energy consumption.

Ⅱ. Methodology

2.1 CityLearn challenge

CityLearn[16] is a simulation environment designed to serve as a testbed for developing and evaluating algorithms related to demand response and energy management in urban settings. Built to mimic real-world scenarios, CityLearn allows for the simulation of various energy management tasks within multiple buildings over extended periods and enables the management of a group of buildings using either centralized or decentralized multi-agent RL control mechanisms, as well as individual-agent RL control systems. Each building has three kinds of demand: electric demand, cooling and DHW demand. Also, buildings include electrical devices such as air-to-water heat pumps, electric heaters, cooling or heating and domestic hot water (DHW) thermal storages, and electrical energy storage, as depicted in Fig 1.

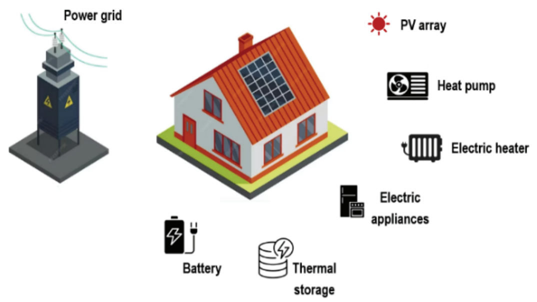

그림 1. 건물의 에너지 모델
Fig. 1 Energy models of buildings

A heat pump is a versatile device that consumes electrical power from the power grid to meet the thermal demand of a building. It operates in two modes: heating and cooling. And these two modes cannot be used at the same time. Electric energy consumed by heat pump is computed as:

$$P_t^{hp} = \frac{Q_t^{hp}}{\eta_t^{hp}} \qquad \cdots (1)$$

$$0 \le Q_t^{hp} \le Q_t^{-\,hp}, \ \ \forall t \in T \qquad \cdots (2)$$

Where η_t^{hp} epresents the energy conversion coefficient, which depends on indoor target and outdoor air temperatures and the technical efficiency coefficient $\eta_{tech}^{hp} Q_t^{-\,hp}$indicates the output power capacity of heat pump.

The electric heater generates heating energy, for Domestic Hot Water (DHW) by utilizing electrical energy from the grid, , in accordance with the following formula:

$$E_{heater} = \frac{Q_{heater}}{\eta_{eh}} \qquad \cdots (3)$$

Where $\eta_{eh}$is heater efficiency and usually greater than 0.9.

Battery capacity, C, is defined in kWh and it includes a capacity loss coefficient, C_{loss} which indicates the fraction of capacity lost during each charge and discharge process. It is defined as $\frac{1}{cycle}$ units. The new battery capacity C_{new} is calculated by following equation:

$$C_{\neq w} = d * \#\ {}_{of_{cycles}} * C_0 \qquad \cdots (4)$$

Where d is capacity degradation rate per cycle. Battery also involves regulating the inflow and outflow of power, which is defined as $P^{ees,in}$ and $P^{ees,out}$respectively.

The storage includes chilled water and domestic hot water (DHW) tanks, which accumulate cooling, heating, and DHW energy supplied by heat pumps and electric heaters. Operation functional of thermal energy storages is analogous to the electric storage system.

The storage includes chilled water and domestic hot water (DHW) tanks, which accumulate cooling, heating, and DHW energy supplied by heat pumps and electric heaters. Operation functional of thermal energy storages is analogous to the electric storage system.

2.2 Observation space

Observation space represents the set of observations the agent perceives at every decision-making interval. Within the CityLearn framework, each building can offer up to 27 distinct observations that can be passed to the agent. Also, CityLearn provides the user with the capability to determine which resources the RL agent can access.

2.3 Action space

In CityLearn environment, agents do not directly control renewable energy sources. Instead, renewable energy is automatically subtracted from the total electricity demand, which is a straightforward approach. However, given our research objective of maximizing renewable energy usage, we decided to grant agents control over solar energy[17-18] generation. This decision enables us to assess the agents' proficiency in managing solar energy. To simplify the system and prevent simultaneous charging and discharging of the battery, we have unified the action space for both charging/discharging the battery and controlling solar generation usage. A positive action (greater than or equal to 0) dictates the proportion of solar generation to be utilized immediately, with the remainder being stored in the battery. Conversely, a negative action (less than 0) signals the battery to discharge, supplementing the energy demand with stored power while using the entirety of the solar generation directly.

It is important to note that the actual amount of energy shifted or stored does not always precisely align with the action value. This discrepancy arises due to the fluctuating nature of building energy demands and the battery's current state of charge. Our approach thus provides a nuanced and practical method for the agents to optimize the use of solar energy in conjunction with battery storage, a crucial step towards achieving our goal of enhancing renewable energy utilization in smart grid environments.

In the context of our study, it is crucial to have periods during the day when there is excess solar generation available for storage. This stored energy can then be utilized during times when solar generation is not possible. Such a scenario is key to evaluating the agents' ability to effectively manage solar energy resources. To determine the most suitable use for solar generation, we conducted an analysis of historical data, comparing historical cooling, domestic hot water (DHW), and electric demand with solar generation data. Our findings indicated that electric demand aligns most effectively with solar generation patterns (Fig. 2).

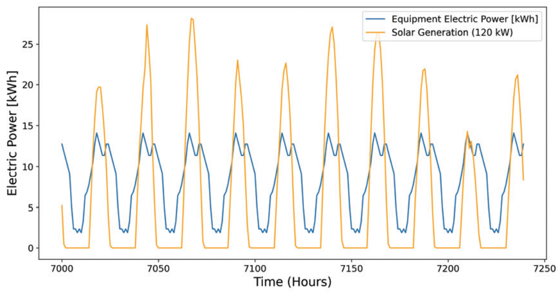

그림 2. 전력 수요와 태양광 발전량 비교
Fig. 2 comparison electric demand and solar generation

2.4 Reward function

We have designed a reward function that integrates the management of solar energy utilization with grid stability. This function comprises two distinct components, each focusing on a specific aspect of energy management.

Reward for solar energy utilization. For each agent, the reward is calculated based on the difference between the non-shiftable demand and the solar energy directly used, as well as the state of charge (SoC) of the battery.

$$\begin{cases} r_1 = \left(d_i * (1 - soc_i)\right)^3 & \text{if } d_i \langle 0 \text{ and } soc_i < 1 \\ r_1 = 0 & \text{if } d_i \langle 0 \text{ and } soc_i = 1 \\ \quad r_1 = -\left(d_i\right)^3 & \quad \text{if } d_i > 0 \end{cases} \quad \cdots (5)$$

Where, represent the difference between non-shiftable load and solar generation for building i, and soc_i represent the state of charge of the battery for building i

Grid Stability Reward. This component focuses on the total electricity demand of each building.

$$r_2 = \left(E_{total}\right)^3 \quad \cdots (6)$$

However, if reward for grid stability is positive, then it is set to 0 to avoid rewarding excessive consumption from the grid. The final for each building is a weighted sum of the two reward components.

2.5 SAC Reinforcement Learning algorithmn

The Soft Actor-Critic (SAC) algorithm is a state-of-the-art reinforcement learning method that is particularly suitable for complex, continuous control tasks, such as those encountered in energy management and optimization in environments. SAC is also known for being sample-efficient.

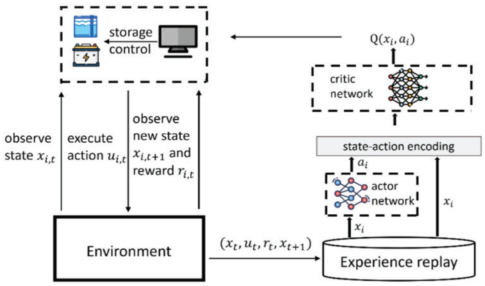

그림 3. SAC 기반 DR 관리를 위한 시스템 설계
Fig. 3 System design for SAC based DR management

This is particularly beneficial in complex simulations like CityLearn, where collecting data can be time-consuming and computationally expensive. The SAC algorithm's balance between exploration and exploitation makes it adaptable to a wide range of tasks and environments, including the diverse scenarios presented in CityLearn (Fig. 3.).

Now, let's briefly go through the architecture and key components of the SAC algorithm:

Actor Network: This network is responsible for policy representation *π(a | s),* mapping states to actions. The policy is typically represented as a neural network with parameters *Θ,* and it outputs a probability distribution over actions given the current state. The objective for the actor network is to maximize the expected return and the entropy of the policy. The policy's objective function can be represented as:

$$J(\pi) = E_{(s_t, a_t) \sim \pi}\left[Q(s_t, a_t) + \alpha H(\pi(\bullet \mid s_t)\right] \cdots (7)$$

Where $Q(s_t, a_t)$ is the action-value function estimated by the critic, $H(\pi(\bullet | s_t)]$ is the temperature parameter that determines the relative importance of the entropy term against the reward, and is the entropy of the policy at state s_t.

Critic Networks: SAC employs two critic networks (twin critics), each parameterized by *ϕ1* and *ϕ2*, which approximate the action-value function *Q(s,a)*. The critics are trained to minimize the Bellman residual:

$$J(Q) = E_{(s_t a_t) \sim D} \left[\begin{array}{l} Q_{\phi_i}(s_t, a_t) - \\ \left(\begin{array}{l} r(s_t, a_t) + \phi E_{s_{t+1} \sim \rho} \\ \left[\begin{array}{l} \min_{i=1,2} Q_{\phi_i}^{'}(s_{t+1}, a_{t+1}^{'}) \\ - \alpha \log \pi(a_t \mid s_t) \end{array} \right] \end{array} \right) \end{array} \right] \cdots (8)$$

Entropy Regularization: Entropy regularization is a key component of SAC, represented by the entropy term $H(\pi(\bullet | s_t)]$ in the actor's objective. It is the expectation of the negative log probability of the action taken according to the policy:

$$H(\pi(\bullet \mid s_t)] = E_{a_t \sim \pi}[\log\pi(a_t \mid s_t)] \quad (9)$$

The temperature parameter adjusts the weighting of this entropy term in the overall objective, balancing exploration and exploitation.

Replay Buffer: SAC uses a replay buffer to store past experiences. The replay buffer stores tuples of experience $(s_t,\ a_t, r_t, s_{t+1})$. It is used to sample mini batches for training, ensuring that the data is uncorrelated and stable.

Soft Updates: The target networks in SAC are updated using a soft updating mechanism, which helps in stabilizing the learning process. This process is controlled by a hyperparameter τ typically a small value close to 0. The update rule for each parameter ϕ_{target} in the target network is:

$$\phi_{target} \leftarrow \tau\phi_{main} + (1-\tau)\phi_{target} \quad \cdots \ (10)$$

Automatic Entropy Tuning: SAC can automatically adjust the weight of the entropy term in its objective function, balancing exploration and exploitation based on the specific demands of the environment.

Ⅲ. Experiments and results

3.1 Data

The platform includes four energy demand datasets, simulated with EnergyPlus for four distinct U.S. climate zones . In this research we use hot-humid climate of Z1, showcasing a year's worth of energy data for a micro-grid consisting of nine buildings, with data recorded every hour. Table 1 illustrates the composition of this building group. As our research goal is maximizing the utilization of renewable energy, we add all buildings all storages and capacity of these storages depends on electricity demands of buildings.

표 1. CityLearn 챌린지의 건물과 설명
Table 1. Building and Descriptions in CityLearn Challenge

Building type	Building description	Cooling Storage	DHW Storage	Battery Capacity(kWh)	PV (kW)
1	Medium Office	3	3	210	120
2	Fast-food restaurant	3	3	160	60
3	Standalone Retail	3	3	160	60
4	Strip Mall	3	3	160	60
5, 6, 7, 8, 9	Multi-Family residental	3	3	120	40

3.2 Simulation parameters

In this experiment, we have implemented a decentralized approach to enhance the efficiency of demand response management. Each building is equipped with individual measurement devices, control systems, and actuators. Consequently, the number of elements in the State, Action, and Reward categories is limited to nine. A notable distinction, however, is that the “net electricity consumption” variable represents the aggregated electricity usage of all nine buildings at a given moment within the State. We can also track solar generation usage by “electric consumption appliances”. This setup enables the trained controller to modify its energy utilization strategy based on the observation of net electricity consumption. The hyperparameters used in training the network are shown in Table 2.

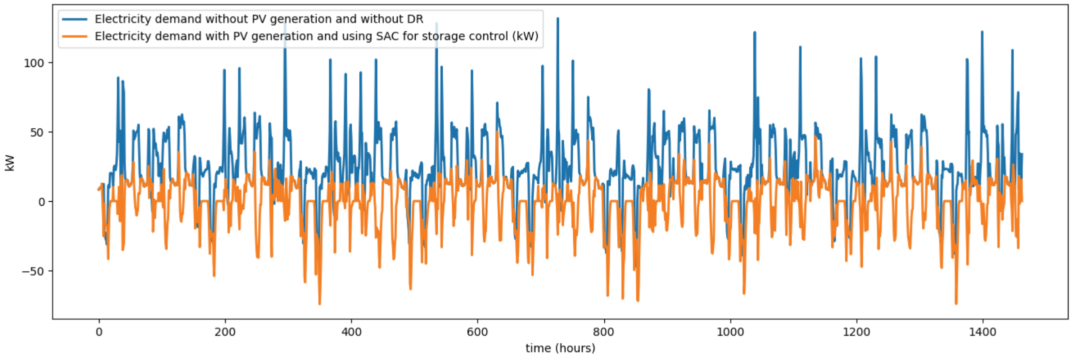

그림 4. 태양광 발전이 있는 전력 수요 및 저장소 및 태양광 발전 제어를 위한 SAC 알고리즘 사용, 태양광 발전 및 DR 관리를 제외한 전력 수요
Fig. 4 Electricity demand with PV generation and using SAC algorithm for storage and solar generation control and Electricity demand without PV generation and DR management

표 2. SAC 알고리즘의 매개변수
Table 2. Parameters of SAC algorithm

Description	Value
Weight for rewards	Battery – 0.6, thermal stor. – 0.4
Discount factor	0.99
Replay buffer	1e5
Mini batch	256
Learning rate	3e-5
τ	5e-4
Learning episodes	20

3.3 Results and analysis

Fig. 4 shows average reward of SAC algorithm per episode. This average reward is for both controlling solar generation and thermal storages. The plot illustrates the average reward per learning episode for the SAC algorithm within the CityLearn simulation environment. Initially, we observe a significant increase in the average reward as the SAC algorithm learns from its interactions with the environment, indicating an improvement in policy and a reduction in associated costs.

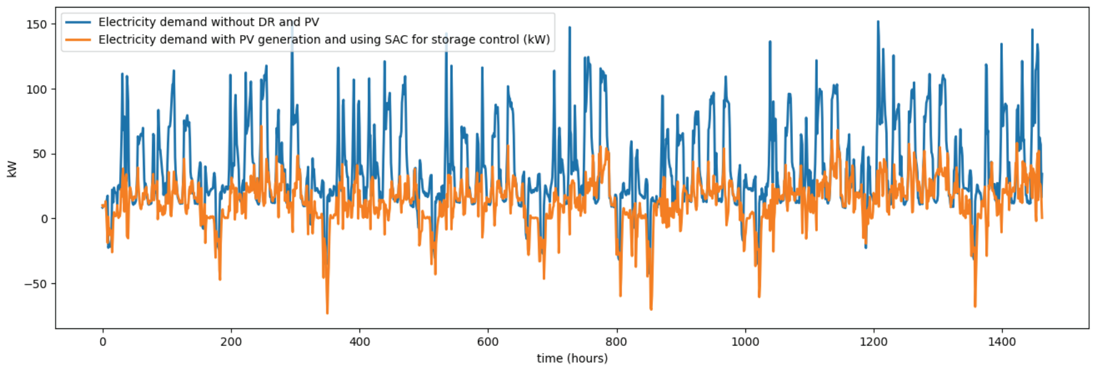

그림 5. 태양광 발전이 있는 총 전력 수요와 저장소 제어를 위해 SAC를 사용하는 총 전력 수요 및 DR 관리 및 태양광 발전이 없는 총 전력 수요
Fig. 5 Total electricity demand with PV generation and using SAC for storage control and total electricity demand without DR management and PV

The reward stabilizes after approximately 5 episodes, suggesting the algorithm has begun to converge towards an optimal policy.

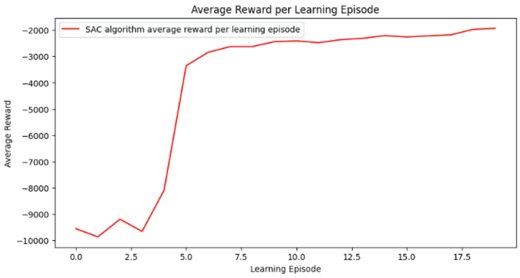

그림 6. SAC 알고리즘 에피소드당 평균 보상
Fig. 6 Average reward SAC algorithm per episode

Fig. 5 and Fig. 6 illustrate total electricity consumption and total electricity appliances consumption using SAC algorithm and storages and without DR and without storages respectively. Notably, the demand profile managed by SAC (orange line) demonstrates a consistent reduction in electricity consumption compared to the baseline scenario devoid of demand response and photovoltaic input (blue line). This reduction is particularly evident during peak demand periods, suggesting that SAC effectively shifts energy usage to off-peak times or utilizes stored solar energy, thereby flattening the demand curve.

Furthermore, instances of negative demand within the SAC-controlled profile indicate moments when solar generation surpasses the building's energy consumption, which not only reduces grid dependency but also provides opportunities for energy storage or export. These observations underscore the SAC algorithm's potential in optimizing demand response actions to leverage renewable energy sources, contributing to a more resilient and sustainable energy system.

Table 3 shows evaluation of SAC algorithm's performance by cost functions. Each cost function assesses a different aspect of the system's efficiency and environmental impact. Ramping measures the rate at which the electricity demand increases or decreases over time. The value close to 1 suggests that the system's performance is almost ideal in minimizing large fluctuations in energy demand. 1-Load Factor is the ratio of the average load over a period to the peak load occurring in that period. In our case, the "1-load factor" being greater than 1 indicates less than ideal performance, with room for improvement in smoothing out energy demand.

표 3. SAC 알고리즘 성능에 대한 평가 지표
Table 3. Evaluation metrics for SAC algorithm performance

Cost function	value
Ramping	0.9960
1-Load Factor	1.0342
Average Daily Peak	0.7846
Peak Demand	0.8415
Net Electricity Consumption	0.8600
Carbon Emissions	0.8651
Total	0.8969

Average Daily Peak averages the highest electricity demand peaks over each day. Peak Demand is the highest recorded electricity demand during the evaluation period. Net Electricity Consumption refers to the total amount of electricity used, accounting for any generation from renewable sources. Carbon Emissions calculates the total carbon emissions resulting from electricity consumption. A lower value indicates a smaller carbon footprint and a more environmentally friendly energy management system. Total is an aggregate score that averages all the other cost function values to give an overall performance metric. The closer this value is to 0, the better the overall performance of the energy management system.

Ⅳ. Key findings

Our research in demand response (DR) using the Soft Actor-Critic (SAC) algorithm within the CityLearn simulation environment yielded significant insights into the optimization of renewable energy consumption and grid stability. The results demonstrate that, the SAC algorithm successfully minimized energy demand volatility, daily peak electricity demand, total electricity consumption and carbon emissions which reflects a lower environmental impact, showing SAC algorithm effectively managing energy sources. The results hold promise for advancing smart grid technologies and contributing to the broader adoption of intelligent energy systems that maximize renewable energy utilization while ensuring grid stability.

Ⅴ. Discussion

The research outcomes from the demand response management using SAC algorithm within the CityLearn simulation environment offer a compelling narrative for the future of smart energy management systems. The SAC algorithm's success in our study corroborates theories suggesting that machine learning can play a pivotal role in advancing the functionality and resilience of smart grids.

Despite the promising results, this research is not without limitations. One such limitation is the occasional drop of non-shiftable load below zero, highlighting an opportunity for further refinement of the algorithm's predictive and adaptive capabilities. This anomaly suggests that the SAC algorithm, while robust, may still require additional tuning to ensure that the non-shiftable loads are consistently predicted and managed within realistic bounds. Future iterations of the algorithm should focus on enhancing load balancing strategies to ensure a consistent and efficient distribution of energy consumption throughout the day.

Ⅵ. Conclusion

Enhancing Smart Grid efficiency through SAC Reinforcement Learning algorithm within the CityLearn environment has led to significant discoveries with substantial implications for the field of demand response and energy management. By focusing solar generation exclusively on battery charging and managing non-shiftable loads, which predominantly consist of electrical appliances, the study has demonstrated an approach to enhancing the efficiency of energy systems while maximizing the use of renewable energy sources. Also, the strategic use of thermal energy storage systems for grid stability has showcased the versatility of the SAC algorithm. This aspect of the research underlines the potential of integrated energy storage solutions in fortifying grid resilience.

The results of the study hold promise for the future of smart energy management. They illustrate the potential of sophisticated reinforcement learning algorithms to not only optimize energy consumption within individual buildings but also to contribute to the broader goal of creating a sustainable and stable energy grid.

감사의 글

This research was supported by "Regional Innovation Strategy (RIS)" through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (MOE)(2021RIS-002)

References

[1] P. Siano, "Demand response and smart grids: A survey," *Renew Sustain Energy Rev.*, vol. 30 no. 37, 2014, pp. 461-478.

[2] J. Joo and J. Oh, "Efficient Grid-Independent ESS control System by Prediction of Energy Production Consumption," *J. of the Korea*

Institute of Electronic Communication Sciences, vol. 14 no. 1, 2020, pp.155-160.

[3] E. Kwak and C. Moon, "Analysis of Power System Stability by Deployment of Renewable Energy Resources," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 16 no. 4, 2021, pp.633-642.

[4] Z. Wang and T. Hong, "Reinforcement learning for building controls: The opportunities and challenges," *Appl Energy*, vol.. 269, 2020. pp.115036

[5] Yang, Shiyu, H. Oliver Gao, and Fengqi You. "Model predictive control for Demand- and MarketResponsive building energy management by leveraging active latent heat storage," *Appl Energy*, vol. 327, 2022. pp.120054

[6] D. Mariano-Hern´andez, L. Hernandez-Callejo, A. Zorita-Lamadrid, O. Duque-P´erez, and F. Santos García, "A review of strategies for building energy management system: Model predictive control, demand side management, optimization, and fault detect & diagnosis," *Journal of Building Engineering*, vol. 33, 2021. pp.101692

[7] C, Henggeler Antunes, M. J. Alves, and I. Soares, "A comprehensive and modular set of appliance operation MILP models for demand response optimization," *Advances Applied Energy*, vol. 320, 2022. pp.119142

[8] J. R. Vazquez-Canteli and Z. Nagy, "Reinforcement learning for demand response: A review of algorithms and modeling techniques," *Advances Applied Energy*, vol. 235 no. 1, 2019, pp. 1072-1089.

[9] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. klimov, "Proximal policy optimization algorithms," *arXiv preprint*, vol. 1707, 2017. pp.06347

[10] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, D. Wiersta, "Continuous control with deep reinforcement learning," in Int. *Conf. on Learning Representations (ICLR)*, San Juan City, PR, USA, 2016, pp. 1-14.

[11] S. Fujimoto, H. V. Hoof, and D. Meger, "Addressing function approximation error in actor-critic methods," *in Proc. of the 35nd Int. Conf. on Machine Learning (ICML)*, Stockholm City, Sweden, 2018, pp. 1587-1596.

[12] A. Ajagekar and F. You, "Deep reinforcement learning based unit commitment scheduling under load and wind power uncertainty," *IEEE Trans Sustain Energy*, vol. 14 no. 2, 2023, pp. 790-803.

[13] R. Lu and Sh. Hong, "Incentive-based demand response for smart grid with reinforcement learning and deep neural network," *Advances Applied Energy*, vol. 236 no. 8, 2019, pp. 937-949.

[14] R. Jin, Y. Zhou, C. Lu, and J. Song, "Deep reinforcement learning-based strategy for charging station participating in demand response," *Advances Applied Energy*, Vol328. 2022, pp.120140

[15] X. Kong, D. Kong, J. Yao, L. Bai, and J. Xiao. "Online pricing of demand response based on long short-term memory and reinforcement learning," *Advances Applied Energy*, vol. 328, 2022. pp.120140

[16] J. V´ azquez-Canteli, S. Dey, G. Henze, and Z. Nagy, "CityLearn: Standardizing Research in Multi-Agent Reinforcement Learning for Demand Response and Urban Energy Management," *arXiv preprint*, vol 2012, 2020. pp.10504

[17] S. Jung, C. Sim, S. Park, and J. Kim, "A Novel of solar Heat collection Device Prototype using Parabolic based on Solar Light Tracking," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 11 no. 3, 2018, pp. 411-420.

[18] M. Kang, "Renewable Energy Generation Prediction Model using Meterological Big Data," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 18 no. 1, 2023, pp. 39-44.

저자 소개

이자노브 알리벡 러스타모비치 (Esanov Alibek Rustamovich)

2022년 전남대학교 대학원 컴퓨터 공학과 졸업(석사)
2023년 ~ 현재 전남대학교 대학원 컴퓨터공학과 재학 (공학박사)
※ 관심분야 : 기계학습, 딥러닝, 빅데이터

성승제 (Seung-Je Seong)

2022년 전남대학교 공학대학 컴퓨터 공학 졸업(학사)
2023년 ~ 현재 전남대학교 대학원 컴퓨터공학과 재학 (공학석사)
※ 관심분야 : 기계학습, 딥러닝, 빅데이터

임창균(Chang-Gyoon Lim)

1997년 Wayne State University, Com puter Engineering(Ph. D.)
2011년~2014 광주테크노파크 가전 로봇 지원센터 센터장
1997년~현재 전남대학교 공학대학 전기·컴퓨터공학부 교수
※ 관심분야 : 인공지능, 빅데이터, 기계학습, 패턴인식, 클라우드 컴퓨팅, 블록체인