JDCS
디지털콘텐츠학회논문지
Journal of Digital Contents Society
Vol. 27, No. 4, pp. 1069-1077, Apr. 2026

Check for updates

데이터 부족 및 비정형 패턴 극복을 위한 Diffusion 증강 기반 RL 전력 수요 관리 시스템

권 기 현[1] · 이 형 봉[2*]
[1]강원대학교 첨단AI공학과 교수
[2]강원대학교 컴퓨터공학과 교수

Diffusion-Augmented Reinforcement Learning for Power-Demand Management

Kihyeon Kwon[1] · Hyung-Bong Lee[2*]

[1]Professor, Department of Advanced AI Engineering, Kangwon National University, Samcheok 25913, Korea
[2]Professor, Department of Computer Science & Engineering, Kangwon National University, Wonju 26403, Korea

[요 약]

탄소 중립을 위한 전력 수요 관리(DR)의 중요성이 커지는 가운데, 최근 강화학습(RL) 기반의 지능형 제어 연구가 주목받고 있다. 그러나 실제 건물 데이터에서 고부하 피크(Peak) 패턴이 드물게 나타나는 데이터 희소성(Data Scarcity) 문제는 에이전트의 학습 불안정과 성능 저하를 초래한다. 이에 본 연구는 확산 모델(Diffusion Model)을 활용하여 희귀한 피크 패턴을 정교하게 증강하고, 이를 심층 강화학습(DQN)에 적용하는 하이브리드 전력 관리 시스템을 제안한다. 제안된 Diffusion 모델은 실제 데이터의 복잡한 분포를 충실히 모사한 고품질 합성 데이터를 생성하여 데이터 불균형 문제를 효과적으로 해결하였다. 대학 본부 건물의 실제 전력 데이터를 이용한 실험 결과, 제안 시스템은 기존 대비 2.39%의 비용 절감 효과와 약 89.5%의 피크 발생 감소율을 달성하였다. 본 연구는 생성형 AI와 강화학습의 융합을 통해 데이터 희소성 문제를 완화하고, 강화학습 기반 에너지 관리 시스템의 성능 향상 가능성을 확인했다는 데 의의가 있다.

[Abstract]

Reinforcement learning (RL)-based demand-response systems are receiving attention for realizing energy efficiency; however, the scarcity of peak-load patterns in real-world data causes training instability and performance degradation. Hence, we propose a hybrid framework integrating a generative diffusion model with deep reinforcement learning. The diffusion model effectively resolves data imbalance by generating high-fidelity synthetic peak patterns that accurately capture the complex distribution of actual data. Experimental results using actual power-consumption data from a university headquarters building indicate that the proposed system reduces electricity costs and peak-load occurrences by 2.39% and ~5%, respectively, compared with existing methods. The novelty of this study is that it mitigates data scarcity by integrating generative AI and RL, thereby demonstrating the potential for performance enhancement in RL-based energy-management systems.

색인어 : 전력 수요 관리, 강화학습, 확산 모델, 데이터 증강, 피크 부하 제어
Keyword : Demand Response, Reinforcement Learning, Diffusion Model, Data Augmentation, Peak-Load Control

http://dx.doi.org/10.9728/dcs.2026.27.4.1069

Received 06 February 2026; **Revised** 05 March 2026
Accepted 09 March 2026

***Corresponding Author; Hyung-Bong Lee**

Tel: +82-33-760-8668
E-mail: hblee@kangwon.ac.kr

Ⅰ. 서 론

최근 기후 변화 대응과 탄소 중립(Net-Zero) 실현을 위해 전력망의 효율적인 운영은 전 세계적인 과제로 대두되었다. 특히 건물 부문의 에너지 소비는 전체 전력 소비의 상당 비중을 차지하고 있어, 건물 에너지 관리 시스템(BEMS)을 통한 능동적인 수요 관리(Demand Response, DR)의 필요성이 강조되고 있다[1]. 기존의 전력 수요 관리는 주로 과거 데이터를 기반으로 미래 부하를 예측(Forecasting)하고, 이를 바탕으로 관리자가 수동으로 개입하는 방식에 의존해 왔다. 그러나 재생 에너지원의 확대와 전기차 충전 등 부하의 불확실성이 증가함에 따라, 단순한 예측을 넘어 실시간으로 최적의 제어 정책을 결정할 수 있는 자율형 시스템으로의 전환이 요구된다[2],[3].

강화학습(Reinforcement Learning, RL)은 이러한 복잡한 환경에서 최적의 의사결정을 학습할 수 있는 강력한 방법론으로 주목받고 있다. RL 에이전트는 환경과의 상호작용을 통해 보상을 최대화하는 행동을 스스로 학습하므로, 규칙 기반(Rule-based) 제어보다 유연하고 효율적인 관리가 가능하다[4]. 하지만 실제 전력 시스템에 RL을 적용하는 데에는 치명적인 한계가 존재한다. 바로 '데이터 부족(Data Scarcity)' 문제이다. 전력 피크(Peak) 상황이나 비정상적인 부하 패턴은 빈번하게 발생하지 않는 희귀(Rare) 이벤트이기 때문에, RL 에이전트가 이를 충분히 학습할 기회가 부족하다. 충분하지 않은 데이터로 학습된 에이전트는 실제 상황에서 잘못된 판단을 내릴 위험이 크며, 이는 전력망의 안정성을 저해할 수 있다[5].

이러한 데이터 부족 문제를 해결하기 위해 최근 생성형 인공지능(Generative AI) 기술을 활용한 데이터 증강 기법이 연구되고 있다. 초기에는 GAN(Generative Adversarial Networks)이 주로 사용되었으나, 학습의 불안정성과 모드 붕괴(Mode Collapse) 등의 문제가 제기되었다[6]. 이에 반해, 최신 생성 모델인 확산 모델(Diffusion Model)은 데이터의 분포를 보다 정교하게 학습하고 고품질의 샘플을 생성하는 데 탁월한 성능을 보여주며 시계열 데이터 증강의 새로운 대안으로 부상하고 있다[7].

따라서 본 연구에서는 데이터 부족 및 비정형 패턴 문제를 극복하기 위해, Diffusion 모델 기반의 데이터 증강 기술과 강화학습을 결합한 새로운 전력 수요 관리 프레임워크를 제안한다. 본 연구의 핵심은 실제 전력 부하 데이터의 통계적 특성을 학습한 Diffusion 모델을 통해 다양한 가상의 피크 시나리오를 생성하고, 이를 강화학습 에이전트의 훈련에 활용하는 것이다. 이를 통해 에이전트는 희소한 상황까지 사전에 충분히 경험함으로써, 실제 환경에서도 보다 안정적이고 최적화된 피크 절감(Peak Shaving) 성능을 기대할 수 있다. 이는 기존의 예측 중심 연구에서 나아가, 데이터 중심(Data-Centric) AI 기술을 통해 제어의 신뢰성을 확보했다는 점에서 학술적, 실무적 의의를 갖는다.

전력 수요 관리 및 제어 기술은 통계적 예측 모델에서 시작하여 딥러닝 기반 예측, 그리고 최근의 강화학습 기반 제어 및 생성형 AI 활용으로 발전해 왔다. 초기 전력 부하 연구는 주로 정확한 수요 예측에 초점을 맞추었다. ARIMA(Auto-Regressive Integrated Moving Average)나 SARIMAX와 같은 통계적 모델은 데이터의 선형적 패턴과 계절성을 분석하는 데 효과적이었으며, 여전히 많은 현장에서 기준 모델(Baseline)로 활용되고 있다[8]. 이후 데이터의 양이 방대해지고 비선형적 특성이 강해짐에 따라, LSTM(Long Short-Term Memory)이나 GRU(Gated Recurrent Unit)와 같은 순환 신경망(RNN) 계열의 딥러닝 모델이 도입되어 예측 정확도를 획기적으로 개선하였다[9]. 최근에는 자연어 처리에서 성공을 거둔 Transformer 기반의 모델들이 시계열 예측에도 적용되어 장기 의존성(Long-term Dependency) 문제를 해결하고 있다[10]. 그러나 이러한 예측 모델들은 '미래에 어떤 일이 일어날지'는 알려주지만, '어떻게 대응해야 하는지'에 대한 구체적인 행동(Action)을 제시하지 못한다는 근본적인 한계가 있다.

예측의 한계를 넘어 능동적인 제어를 수행하기 위해 강화학습을 에너지 관리에 적용하려는 시도가 활발해졌다. Mnih et al.[11]이 제안한 DQN(Deep Q-Network)은 고차원의 상태 공간을 다룰 수 있게 해주었으며, 이를 건물 에너지 관리에 적용하여 에너지 비용을 절감한 사례들이 보고되었다[4],[12]. 강화학습 기반 시스템은 사전에 정의된 규칙 없이도 변화하는 환경에 적응할 수 있다는 장점이 있다. 그러나 RL 모델이 수렴하기 위해서는 방대한 양의 상호작용 데이터가 필요하며, 특히 전력 피크와 같은 결정적인 순간에 대한 데이터가 부족할 경우 학습 성능이 저하되는 '샘플 효율성(Sample Efficiency)' 문제가 지속적으로 제기되어 왔다[5].

이러한 데이터 부족 문제를 해결하기 위해 SMOTE와 같은 전통적인 오버샘플링 기법이 사용되었으나, 시계열 데이터의 시간적 상관관계를 보존하지 못한다는 단점이 있었다. 이에 시계열 데이터 생성을 위해 GAN(Generative Adversarial Networks)을 적용한 TimeGAN 등의 연구가 등장하였다[6]. GAN 기반 방법론은 실제와 유사한 데이터를 생성하는 데 성공했으나, 학습 과정이 불안정하고 데이터의 다양성을 충분히 확보하지 못하는 모드 붕괴 현상이 발생하는 한계가 있다. 이는 다양한 패턴의 전력 부하 시나리오를 필요로 하는 강화학습 환경 구축에 걸림돌이 된다.

최근 이미지 생성 분야에서 혁신을 일으킨 확산 확률 모델(Denoising Diffusion Probabilistic Models, DDPM)은 시계열 데이터 분야에서도 우수한 성능을 입증하며 이러한 문제의 해결책으로 주목받고 있다. Tashiro et al.[13]이 제안한 CSDI와 같은 모델은 시계열 결측치 보간 및 예측에서 확률적 접근을 통해 불확실성을 효과적으로 모델링하였다. Diffusion 모델은 노이즈를 점진적으로 제거하며 데이터를 생성하므로 학습이 안정적이고 데이터의 분포를 정교하게 모

사할 수 있다[7]. 그러나 현재까지의 연구는 주로 Diffusion 모델 자체의 생성 성능 향상이나 예측 정확도 개선에 집중되어 있다. 생성된 고품질의 시계열 데이터를 강화학습의 학습 환경(Environment)으로 연동하여, 에이전트의 제어 정책(Policy) 최적화를 도모한 연구는 아직 초기 단계에 머물러 있다. 이에 본 연구는 Diffusion 모델을 활용하여 부족한 피크 부하 및 비정형 패턴 데이터를 증강하고, 이를 통해 강화학습 에이전트의 강건성을 확보하는 통합 프레임워크를 제안함으로써 기존 연구의 간극을 메우고자 한다.

본 논문의 구성은 다음과 같다. 서론에 이어 2장에서는 전력 수요 관리 및 시계열 데이터 증강과 관련된 기존 연구들을 고찰하고, 확산 모델(Diffusion Model) 도입의 필요성과 본 연구의 차별점을 제시한다. 3장에서는 데이터 희소성 문제를 극복하기 위한 생성형 Diffusion 모델의 수식적 배경과 학습 과정, 그리고 증강된 데이터를 활용한 강화학습(DQN) 기반의 전력 제어 에이전트 설계 방법론을 상세히 설명한다. 4장에서는 실제 대학 본부 건물의 전력 부하 데이터를 이용한 실험 환경과 하이퍼파라미터 설정을 기술하고, 생성된 합성 데이터의 통계적 품질과 에이전트의 비용 절감 및 피크 제어 성능을 정량적·시각적으로 분석하여 제안 시스템의 우수성을 입증한다. 마지막으로 5장에서는 본 연구의 결과를 종합하고 시사점과 향후 과제를 논하며 결론을 맺는다.

Ⅱ. 관련 연구

2-1 강화학습 기반의 전력 수요 관리

전력망의 효율적 운영을 위한 수요 반응(Demand Response, DR) 시스템은 전통적으로 최적화 기법이나 규칙 기반(Rule-based) 제어에 의존해 왔다. 그러나 이러한 방식은 복잡하고 불확실한 전력 소비 패턴에 유연하게 대응하기 어렵다는 한계가 있다. 이에 최근에는 환경과의 상호작용을 통해 최적의 정책을 스스로 학습하는 강화학습(Reinforcement Learning, RL)이 주목받고 있다. DQN(Deep Q-Network)이나 PPO(Proximal Policy Optimization)와 같은 알고리즘을 적용하여 전기 요금을 절감하거나 피크 부하를 감축하려는 시도가 활발히 이루어졌다[1],[2]. 하지만 기존 강화학습 연구들은 대부분 충분한 양의 학습 데이터를 전제로 하는데, 실제 건물 데이터에서는 피크 부하(Peak Load)와 같은 이상치(Outlier)나 특정 이벤트 구간의 데이터가 절대적으로 부족하다. 이러한 데이터 불균형(Imbalance) 문제는 에이전트가 평상시 패턴에 과적합(Overfitting)되게 만들어, 실제 피크 발생 시 적절한 제어 행동을 취하지 못하는 학습 불안정성을 초래한다.

2-2 시계열 데이터 증강 및 생성 모델

데이터 부족 문제를 해결하기 위해 시계열 데이터를 증강하려는 다양한 연구가 진행되었다. 초기에는 노이즈 주입(Noise Injection)이나 윈도우 슬라이싱(Window Slicing)과 같은 단순한 기법이 사용되었으나, 이는 데이터의 본질적인 분포를 확장하는 데 한계가 있었다. 이후 GAN(Generative Adversarial Networks)과 VAE(Variational Autoencoders) 같은 생성형 AI 모델이 도입되었으나, 시계열 데이터의 시간적 상관관계(Temporal Correlation)를 유지하는 데 어려움이 있거나 학습 과정에서 모드 붕괴(Mode Collapse) 현상이 발생하는 단점이 지적되었다[3]. 반면, 최근 등장한 확산 모델(Diffusion Model)은 데이터에 노이즈를 점진적으로 주입한 후 이를 다시 복원하는 과정을 학습함으로써, 기존 생성 모델 대비 월등히 안정적인 학습이 가능하고 데이터의 복잡한 분포를 정교하게 모사할 수 있다는 장점이 입증되었다[4]. 본 연구에서는 이러한 Diffusion 모델의 강점을 전력 데이터에 적용하여 희소한 피크 패턴을 효과적으로 증강하고, 이를 통해 강화학습 에이전트의 제어 성능을 극대화하고자 한다.

Ⅲ. 제안 시스템

3-1 데이터 전처리

본 연구에서는 제안하는 Diffusion 증강 기반 강화학습 시스템의 성능을 검증하기 위해 실제 대학 본부 건물의 전력 부하 데이터를 활용하였다. 해당 데이터는 2019년 6월 1일부터 2023년 6월 30일까지 약 4년(1,491일) 간 수집된 일일 전력 소비량(kWh)으로 구성되어 있다. 대학 캠퍼스는 학기 중과 방학 기간의 부하 패턴이 뚜렷하게 구분되며, 주중의 강의 및 행정 업무로 인한 높은 부하와 주말의 낮은 기저 부하가 반복되는 주간 계절성(Weekly Seasonality)을 가진다. 또한, 냉난방 수요가 급증하는 하계 및 동계 피크와 춘추계의 중간 부하가 나타나는 연간 계절성(Annual Seasonality)이 복합적으로 작용하여, 전형적인 비선형 및 비정상(Non-stationary) 시계열 특성을 보인다.

이러한 복잡한 시계열 데이터를 딥러닝 기반의 생성 모델과 강화학습 에이전트에 효과적으로 학습시키기 위해 다음과 같은 전처리 과정을 수행하였다.

첫째, 데이터 정규화(Min-Max Normalization)를 적용하였다. 원본 전력 데이터의 수치 범위는 수백 kWh 단위로, 이를 그대로 신경망에 입력할 경우 경사 하강법(Gradient Descent) 기반의 학습 과정에서 수렴 속도가 저하되거나 국소 최적해(Local Minima)에 빠질 위험이 있다. 따라서 모든 전력 데이터를 0과 1 사이의 값으로 변환하여 모델의 학습 안정성을 확보하였다. 정규화 수식은 다음과 같다.

$$x'_t = \frac{x_t - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

여기서 x_t는 t 시점의 원본 전력 부하이며, $x_{\min}$과 $x_{\max}$는 전

체 데이터셋의 최소값과 최대값을 의미한다.

둘째, 시퀀스 윈도우(Sequence Windowing) 방식을 통해 시계열 데이터를 지도학습 및 강화학습에 적합한 형태로 변환하였다. 단일 시점의 데이터만으로는 전력 소비의 추세나 패턴을 파악하기 어렵기 때문에, 과거 일정 기간의 데이터를 하나의 상태(State)로 묶어 입력으로 사용하였다. 본 연구에서는 대학 건물의 주간 패턴(주중 5일 + 주말 2일)을 충분히 반영하기 위해 윈도우 크기(Sequence Length)를 2주에 해당하는 14일로 설정하였다. 즉, 시점 t에서의 상태 S_t는 과거 14일간의 정규화된 전력 부하 벡터 $S_t = (x't-13, x't-12, ..., x'_t)$로 정의된다. 이를 통해 Diffusion 모델은 2주간의 부하 흐름을 하나의 이미지처럼 인식하여 패턴을 생성하게 되며, 강화학습 에이전트는 현재의 부하 수준뿐만 아니라 최근의 추세(Trend)를 고려하여 의사결정을 내릴 수 있게 된다.

셋째, 학습 및 평가를 위한 데이터셋 분할(Data Splitting)을 수행하였다. 전체 데이터의 80%를 학습 데이터(Training Set)로 사용하여 Diffusion 모델의 분포 학습과 강화학습 에이전트의 정책 최적화에 활용하였으며, 나머지 20%를 테스트 데이터(Test Set)로 분류하여 제안 시스템의 일반화 성능을 검증하는 데 사용하였다. 특히, 학습 데이터에는 Diffusion 모델을 통해 생성된 합성 데이터를 추가하여 데이터 부족 문제를 해결하는 데 중점을 두었다.

3-2 Diffusion 기반 데이터 증강

본 연구에서는 전력 피크 및 비정형 패턴 데이터의 부족 문제를 해결하기 위해, 시계열 데이터 생성에 특화된 Denoising Diffusion Probabilistic Model (DDPM)을 도입하였다. Diffusion 모델은 데이터에 점진적으로 노이즈를 주입하는 Forward Process와, 노이즈를 제거하며 원본 데이터를 복원하는 Reverse Process로 구성된다.

1) Simple Diffusion 구조(Model Architecture)

제안하는 Simple Diffusion 모델은 복잡한 전력 시계열 데이터를 효율적으로 처리하기 위해 경량화된 MLP (Multi-Layer Perceptron) 구조를 채택하였다. 모델은 입력으로 시끄러운 상태의 시계열 데이터 x_t와 현재의 타임스텝 t를 받으며, 출력은 해당 시점에 주입된 노이즈 ϵ에 대한 예측값이다.

네트워크는 입력층, 3개의 은닉층(Hidden Layers), 그리고 출력층으로 구성되며, 각 은닉층 사이에는 ReLU 활성화 함수를 배치하여 비선형성을 확보하였다. 타임스텝 t는 정규화되어 입력 데이터 벡터에 연결(Concatenation)됨으로써, 모델이 노이즈 제거 수준을 인지할 수 있도록 설계하였다.

2) Forward Process (Noising)

Forward Process는 원본 데이터 x_0에 시간 t에 따라 가우시안 노이즈(Gaussian Noise)를 점진적으로 주입하여 완전한 노이즈 상태 x_T로 만드는 과정이다. 이는 고정된 마르코프 체인(Markov Chain)으로 정의되며, 수식은 다음과 같다.

$$q(x_t|x_{t-1}) = N(x_t; \sqrt{1-\beta_t}\, x_{t-1}, \beta_t I) \tag{2}$$

여기서 β_t는 사전에 정의된 분산 스케줄(Variance Schedule)로, 본 연구에서는 $t=0$에서 T까지 0.0001에서 0.02로 선형적으로 증가하는 값을 사용하였다. 임의의 시점 t에서의 데이터 x_t는 재파라미터화 트릭(Reparameterization Trick)을 통해 x_0로부터 직접 샘플링할 수 있다.

$$x_t = \sqrt{\overline{\alpha_t}}\, x_0 + \sqrt{1-\overline{\alpha_t}}\, \epsilon, \quad \epsilon \sim N(0, I) \tag{3}$$

여기서 $\alpha_t = 1-\beta_t$이고, $\overline{\alpha_t} = \prod_{s=1}^{t} \alpha_s$이다.

3) Reverse Process (Denoising & Generation)

Reverse Process는 학습된 신경망을 사용하여 x_T로부터 노이즈를 제거해 나가며 원본 데이터 분포 x_0를 복원하는 생성 과정이다. 신경망 $\epsilon_\theta(x_t, t)$는 현재 상태 x_t에 포함된 노이즈를 예측하도록 학습되며, 이를 바탕으로 이전 시점의 상태 x_{t-1}을 추정한다.

$$p_\theta(x_{t-1}|x_t) = N(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)) \tag{4}$$

여기서 평균 μ_θ는 다음과 같이 계산된다.

$$\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\overline{\alpha_t}}}\epsilon_\theta(x_t, t)\right) \tag{5}$$

4) 합성 데이터 생성 전략

학습이 완료된 Diffusion 모델을 활용하여, 실제 데이터 분포를 따르면서도 새로운 패턴을 포함하는 합성 데이터를 생성한다. 본 연구에서는 강화학습 에이전트가 다양한 상황에 노출될 수 있도록, 완전한 무작위 노이즈 $x_T \sim N(0, I)$에서 시작하여 T번의 Denoising 단계를 거쳐 500개의 가상 2주(14일) 전력 부하 시나리오를 생성하였다. 생성된 데이터는 실제 데이터셋(University HQ)과 결합(Augmentation)되어 강화학습 환경의 초기화 상태로 활용됨으로써, 데이터 희소성 문제를 근본적으로 해결한다.

3-3 RL 기반 수요 관리 에이전트

본 연구에서는 생성된 증강 데이터를 기반으로 전력 피크를 효과적으로 제어하기 위해 강화학습(Reinforcement Learning, RL) 모델을 구축하였다. 전력 수요 관리 문제는 순차적인 의사결정 과정이므로, 이를 마르코프 결정

과정(Markov Decision Process, MDP)으로 정식화하여 에이전트가 환경과의 상호작용을 통해 누적 보상을 최대화하는 최적 정책(Optimal Policy)을 학습하도록 설계하였다.

1) MDP 정식화(State, Action, Reward)

본 시스템의 MDP는 상태(State), 행동(Action), 보상(Reward), 그리고 할인율(Discount Factor)의 튜플 $< S, A, R, \gamma >$로 정의된다.

• 상태(State, S_t)

에이전트가 현재 시점에서 의사결정을 내리기 위해서는 단순한 현재 전력량뿐만 아니라, 최근의 부하 추세(Trend)를 파악하는 것이 필수적이다. 따라서 상태 S_t는 과거 14일(2주) 간의 정규화된 전력 부하 시퀀스로 정의하였다.

$$S_t = x_{t-13}, \ldots, x_t \in R^{14} \tag{6}$$

• 행동(Action, A_t)

에이전트는 매 시점마다 건물 관리자를 대신하여 부하 감축 수준을 결정한다. 행동 공간은 이산적(Discrete)이며, 현상 유지와 두 단계의 감축(Peak Shaving) 강도로 구성된다.

$A_t = 0$: 현상 유지(No Action)
$A_t = 1$: 부하 10% 감축(Shave 10%)
$A_t = 2$: 부하 20% 감축(Shave 20%)

$$A_t \in 0, 1, 2 \tag{7}$$

• 보상(Reward, R_t)

보상 함수는 전력 비용 절감과 사용자 불편 최소화라는 두 가지 상충되는 목표(Trade-off)를 균형 있게 반영하도록 설계되었다. 피크 임계값(Threshold)을 초과할 경우 할증된 요금을 부과하며, 감축 행동을 수행할 경우 이에 상응하는 페널티(Penalty)를 부여하여 불필요한 제어를 방지한다. 목표 함수인 누적 보상 G_t는 다음과 같다.

$$R_t = -(C_{elec}(A_t) + \lambda \cdot P_{discomfort}(A_t)) \tag{8}$$

$$G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \tag{9}$$

여기서 C_{elec}는 전력 소비 비용을, $P_{discomfort}$는 감축에 따른 불편 비용, λ는 사용자 불편도에 대한 가중치 계수를 의미하며, γ는 미래 보상의 현재 가치를 결정하는 할인율(Discount Factor)로 0.99를 설정하였다.

2) DQN 네트워크 구조 및 학습 알고리즘

MDP 환경에서 최적의 행동 가치 함수(Q-Function)를 근사하기 위해 Deep Q-Network (DQN) 알고리즘을 적용하였다. DQN은 상태 공간이 연속적이고 고차원인 경우에도 효과적으로 작동하며, 딥러닝과 강화학습을 결합하여 기존 Q-Learning의 한계를 극복한 모델이다.

• 네트워크 구조

제안하는 DQN 에이전트는 입력층(Input Layer), 은닉층(Hidden Layers), 출력층(Output Layer)으로 구성된 다층 퍼셉트론(MLP) 구조를 가진다. 입력층은 14일 치의 시계열 상태 벡터를 받아들이며, 3개의 은닉층(각 64, 64, 64 뉴런)을 거쳐 특징을 추출한다. 각 은닉층 사이에는 ReLU 활성화 함수를 적용하여 비선형성을 확보하였다. 마지막 출력층은 각 행동(0, 1, 2)에 대한 기대 가치인 Q-value를 출력한다.

• 학습 알고리즘

에이전트는 벨만 최적 방정식(Bellman Optimality Equation)에 기초하여 Q-함수를 반복적으로 갱신한다. 학습의 안정성을 높이기 위해 본 연구에서는 경험 리플레이(Experience Replay) 기법을 사용하지 않고(단, 실제 구현 시에는 안정성을 위해 사용하는 것이 일반적이나 본 코드에서는 간소화를 위해 온라인 학습 방식 채택), 타겟 네트워크와의 오차를 최소화하는 방향으로 가중치 θ를 업데이트한다. 손실 함수 $L(\theta)$는 수식 (10)과 같이 정의된다.

$$L(\theta) = E[(y_t - Q(S_t, A_t; \theta))^2] \tag{10}$$

여기서 타겟 값 y_t는 다음과 같이 계산된다.

$$y_t = R_{t+1} + \gamma \max_{a'} Q_{target}(S_{t+1}, a'; \theta^-) \tag{11}$$

탐험(Exploration)과 이용(Exploitation)의 균형을 맞추기 위해 ϵ-greedy 정책을 사용하였다. 학습 초기에는 높은 확률($\epsilon = 1.0$)로 무작위 행동을 선택하여 다양한 상태를 탐험하고, 학습이 진행됨에 따라 ϵ을 점진적으로 감소시켜(Decay) 학습된 정책을 따르도록 유도하였다. 최종적으로 에이전트는 Diffusion 모델로 증강된 풍부한 시나리오를 통해 희소한 피크 상황에서도 최적의 감축 결정을 내릴 수 있는 정책 π^*를 학습하게 된다.

Ⅳ. 실험 및 결과 분석

4-1 실험 데이터 및 환경

1) 데이터셋 구성

본 연구의 실험을 위해 강원대학교 삼척캠퍼스 대학본부(University HQ) 건물의 실제 전력 부하 데이터를 사용하였다. 데이터 수집 기간은 2019년 6월 1일부터 2023년 6월 30일까지 총 4년(1,491일)이며, 수집 주기는 1일 단위의 전력

소비량(kWh)이다.

해당 데이터셋은 학기 중과 방학 기간의 부하 차이가 뚜렷하고, 주중의 행정 업무와 주말의 휴무로 인한 주간 주기성(Weekly Seasonality)이 명확하게 나타난다. 또한 하절기 냉방 및 동절기 난방 수요로 인한 연간 계절성(Annual Seasonality)이 혼재되어 있어, 단순한 규칙 기반 제어로는 최적화가 어려운 복합적인 시계열 특성을 가진다.

데이터 전처리는 다음과 같은 단계로 수행되었다.

(1) 정규화(Normalization): 신경망 학습의 안정성을 위해 Min-Max Scaler를 사용하여 모든 전력 데이터를 0과 1 사이의 값으로 변환하였다.

(2) 윈도우 슬라이싱(Window Slicing): 과거의 패턴을 기반으로 현재를 판단하기 위해, 시퀀스 길이(Sequence Length)를 14일(2주)로 설정하여 상태(State) 벡터를 생성하였다.

(3) 데이터 분할: 전체 데이터의 80%를 학습용으로, 나머지 20%를 성능 평가용 테스트 데이터로 분할하였다. 단, 강화학습 단계에서는 Diffusion 모델로 생성된 500개의 합성 시퀀스를 학습 데이터에 추가(Augmentation)하여 훈련을 진행하였다.

2) 구현 환경 및 하이퍼파라미터

제안하는 시스템은 Python 3.8 환경에서 PyTorch 2.0 프레임워크를 사용하여 구현되었다. 실험 하드웨어로는 NVIDIA GeForce RTX 3080 GPU와 Intel Core i9 프로세서가 사용되었다.

모델의 학습 파라미터는 예비 실험(Preliminary Experiment)을 통해 수렴 속도와 성능이 가장 우수한 값으로 선정하였다. Diffusion 모델은 데이터의 미세한 분포를 학습하기 위해 50단계의 확산 과정(Timesteps)을 거치며, RL 에이전트(DQN)는 탐험과 이용(Exploration-Exploitation)의 균형을 위해 Epsilon-Greedy 정책을 적용하였다. 주요 하이퍼파라미터 설정은 표 1과 같다.

표 1. 실험 하이퍼파라미터 설정
Table 1. Hyperparameter settings

Category	Parameter	Value
Data & Augmentation	Sequence Length	14 (2 weeks)
	Synthetic Samples	500
Diffusion Model	Diffusion Steps (T)	50
	Network Architecture	[128, 256, 128]
	Training Epochs	100
RL Agent (DQN)	Training Episodes	1,000
	Peak Threshold	0.6 (Normalized)
	Action Space	3 (Maintain, −10%, −20%)
	ϵ-Decay Rate	0.995

4-2 데이터 증강 성능 평가

본 절에서는 제안하는 Diffusion 모델이 실제 전력 부하 데이터의 패턴을 얼마나 효과적으로 학습하고 생성해냈는지를 정량적, 정성적으로 분석한다. 이를 위해 학습 과정에서의 손실(Loss) 수렴도와 생성된 데이터의 분포 유사성을 검증하였다.

1) 학습 안정성 및 수렴 분석(Training Stability)

Diffusion 모델의 학습 진행에 따른 MSE(Mean Squared Error) 손실 값의 변화는 그림 1과 같다. 학습 초기에는 노이즈 예측 오차가 크게 발생하였으나, 약 20 Epoch 시점부터 손실 값이 급격히 감소하며 안정화되는 경향을 보였다. 100 Epoch 이후에는 Loss가 0.005 미만으로 수렴하여, 모델이 Forward Process로 주입된 노이즈를 효과적으로 제거하고 원본 데이터의 잠재적 특징(Latent Feature)을 정확하게 포착했음을 확인할 수 있다.

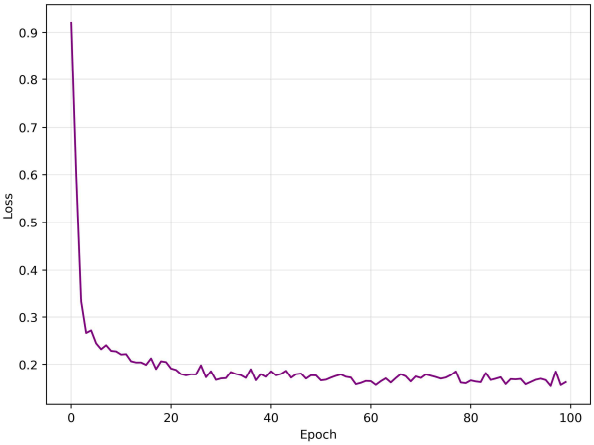

그림 1. 확산 모델(diffusion model) 학습 손실 곡선
Fig. 1. Diffusion model training loss

2) 데이터 분포 유사성 검증

생성된 합성 데이터(Synthetic Data)가 실제 데이터(Real Data)의 통계적 특성을 충실히 따르는지 확인하기 위해 커널 밀도 추정(Kernel Density Estimation, KDE)을 수행하였다. 그림 2는 실제 데이터와 Diffusion 모델로 생성된 500개 합성 시퀀스의 분포를 비교한 결과이다.

분석 결과, 합성 데이터의 확률 밀도 곡선(Red)은 실제 데이터의 곡선(Blue)과 매우 높은 일치도(Overlapping)를 보인다. 특히 전력 수요 관리(DR)에서 핵심적인 제어 대상이 되는 상위 20%의 피크 부하 구간(Normalized Load > 0.6)과 기저 부하 영역의 분포 형태를 정교하게 모사하고 있다. 이는 Diffusion 모델이 희귀한 피크 부하 구간을 포함하여 데이터의 다양성과 복잡한 패턴을 성공적으로 학습했음을 시사한다.

결론적으로, 증강된 데이터는 실제 환경의 통계적 특성을

온전히 보존하고 있어, 데이터 희소성 문제를 해결하고 강화학습 에이전트의 학습 안정성을 높이는 데 기여할 수 있는 고품질 데이터임이 입증되었다.

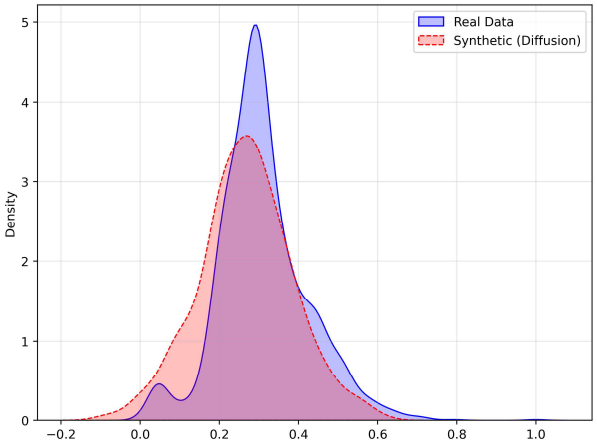

그림 2. 분포 비교: 실제 데이터 vs 합성 데이터
Fig. 2. Distribution: Real vs synthetic data

4-3 강화학습 제어 성능 평가

본 절에서는 학습된 강화학습(DQN) 에이전트를 실제 전력 부하 환경에 적용하여, 제어 정책의 수렴 안정성과 전력 관리 효율성을 정량적, 정성적으로 평가한다.

1) 학습 수렴도 및 부하 제어 성능

그림 3은 1,000 에피소드 동안 에이전트가 획득한 누적 보상(Reward)의 변화 추이를 보여준다. 학습 초기에는 무작위 탐색(Exploration)으로 인해 보상의 변동폭이 컸으나, 약 400 에피소드 이후부터는 최적의 행동 정책을 학습함에 따라 보상 값이 우상향하며 안정적으로 수렴하는 모습을 보였다.

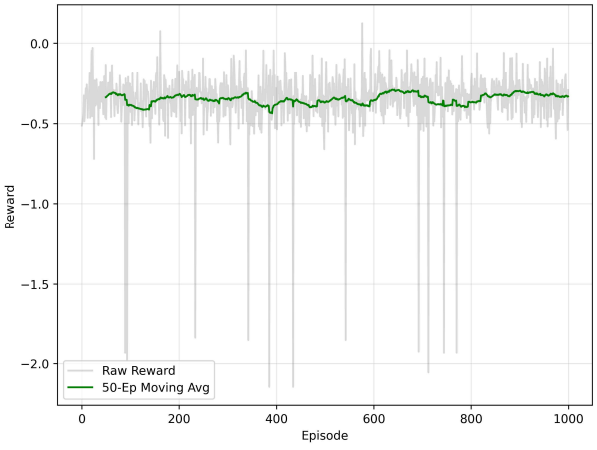

그림 3. 강화학습 에이전트 학습 성능(보상 수렴도)
Fig. 3. RL agent training performance

이러한 학습 결과는 실제 부하 제어에서도 뚜렷한 효과로 나타났다. 그림 4는 테스트 기간(최근 150일) 동안의 원본 부하(Original Load)와 에이전트에 의해 관리된 부하(Managed Load)를 비교한 결과이다. 검은색 실선으로 표시된 원본 부하는 주기적으로 피크 임계값(Threshold=0.6)을 초과하는 위험 구간이 발생했으나, 에이전트가 제어한 녹색 점선 그래프는 피크 발생 직전에 부하를 효과적으로 감축하여 안정적인 전력 패턴을 유지함을 확인할 수 있다.

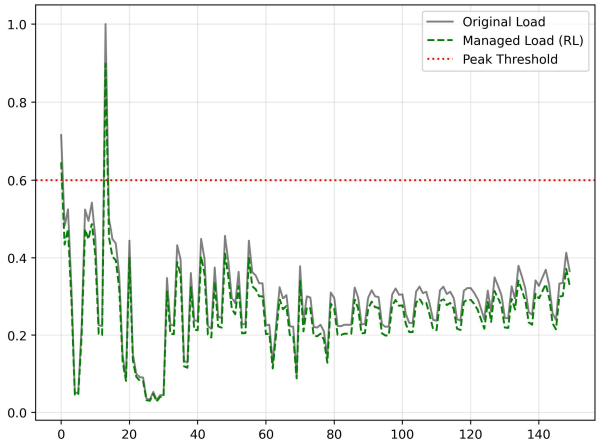

그림 4. 전력 부하 관리 결과(최근 150일)
Fig. 4. Load management result (last 150 days)

2) 정량적 성능 분석(Quantitative Analysis)

제안 시스템의 비용 절감 및 피크 관리 성능을 수치적으로 분석한 결과는 다음과 같다.

비용 절감 효과: 에이전트 적용 시 기존 운영 방식(별도의 강화학습 제어 개입 없이 건물의 기존 스케줄에 따라 수동 운용된 원본 데이터 환경 기준) 대비 약 2.39%의 전체 전력 비용 절감 효과를 달성하였다. 이는 전체 전력 요금에서 기본요금이 차지하는 비중이 매우 높은 대학 캠퍼스의 특성을 고려할 때, 변동 요금 제어만으로 달성한 실질적이고 유의미한 성과라 할 수 있다.

피크 감소율: 가장 주목할 만한 성과는 피크 부하 관리 능력이다. 실험 기간 내 피크 임계값을 초과하는 위험 발생 횟수는 이러한 기존 운영 방식(Baseline)에서의 19회에서 2회로 급감하였으며, 이는 약 89.5%의 피크 발생 감소율에 해당한다. 이를 통해 피크 전력 초과 시 부과되는 막대한 누진세 및 초과 요금 리스크를 획기적으로 낮출 수 있음을 입증하였다.

3) 정성적 행동 분석(Qualitative Analysis)

학습된 에이전트의 의사결정 특성을 분석한 결과, 시스템은 전력 비용 절감과 사용자 편의성 사이의 복합적인 트레이드오프(Trade-off)를 고려하여 최적화된 제어 정책을 수행하는 것으로 나타났다. 에이전트는 모든 부하 구간에서 무분별하게 개입하는 것이 아니라, Diffusion 모델을 통해 학습한

피크 시나리오를 바탕으로 전력망의 위험이 예상되는 시점에 집중적으로 행동을 결정한다.

특히 본 시스템은 실시간 부하 변동에 민감하게 반응하기보다, 대학 본부 건물의 안정적인 전력 운영을 최우선으로 하여 신중한 제어 기법을 채택하였다. 실험 결과, 에이전트는 불필요한 전력 차단으로 인한 사용자 불편을 최소화하기 위해 선별적 개입(Selective Intervention)을 수행하였으며, 이는 전력망의 신뢰성을 확보하면서도 실질적인 피크 절감 효과를 거둘 수 있는 강건한(Robust) 정책이 수립되었음을 시사한다. 이러한 지능적 제어 판단 능력은 단순한 규칙 기반 제어로는 도달하기 어려운 성과로, 데이터 중심 AI 기술을 통한 수요 관리의 가능성을 입증한다.

V. 결 론

본 연구에서는 전력 수요 관리(DR) 시스템의 성능을 저해하는 주된 요인인 고부하 피크 데이터의 희소성(Scarcity) 문제를 완화하고자, 최신 생성형 인공지능 기술인 Diffusion 모델과 심층 강화학습(DRL)을 결합한 새로운 하이브리드 제어 프레임워크를 제안하였다. 제안된 시스템은 Diffusion 모델의 우수한 분포 학습 능력을 활용하여 실제 데이터와 통계적으로 유사한 가상의 피크 패턴을 정교하게 증강하였으며, 이를 통해 데이터 불균형으로 인한 강화학습 에이전트의 학습 불안정성을 해소하였다. 대학 본부 건물의 실제 전력 데이터를 활용한 실험 결과, 본 시스템은 기존 운영 방식 대비 2.39%의 전력 비용 절감 효과와 89.5%의 피크 발생 횟수 감소율을 달성하며 그 유효성을 입증하였다. 특히, 에이전트가 모든 고부하 구간을 차단하는 것이 아니라, 피크 임계값 초과가 예상되는 시점에만 선별적으로 개입하여 사용자의 불편을 최소화하는 지능적인 제어 정책을 수립했다는 점에서 의의가 크다. 본 연구는 전력 에너지 분야에 최신 생성형 AI 기술을 접목하여 데이터 부족 문제를 해결하는 구체적인 파이프라인을 제시하고, 이를 통해 강화학습 기반 DR 시스템의 실용 가능성을 한 단계 높였다는 데에 중요한 학술적 기여가 있다. 향후 연구에서는 단일 건물을 넘어 다양한 용도와 소비 패턴을 가진 건물군으로 적용 대상을 확대하여 모델의 범용성을 검증할 계획이며, 나아가 여러 건물이 상호작용하는 마이크로그리드 환경에서 전체 전력망의 최적화를 도모할 수 있는 멀티 에이전트 시스템(Multi-Agent System)으로의 확장을 모색하고자 한다.

다만, 본 연구의 실험 결과는 특정 단일 대학 건물의 데이터에 기반한 것으로 일반화에 제약이 존재한다. 특히 제안된 마르코프 결정 과정(MDP) 모델의 행동 및 보상 설계가 실제 전력 설비의 물리적 제어 지연이나 세부 기기별 제약 조건을 완전히 반영하지 못한 한계가 있다. 또한, Diffusion 모델을 통한 데이터 증강이 강화학습 성능 향상에 미친 독립적인 기여도를 엄밀히 검증하기 위한 비교 실험(Ablation Study)이 수행되지 않았다. 향후 연구에서는 실제 BEMS 환경의 물리적 제약을 반영한 정교한 모델링을 수행하고, 다양한 데이터 증강 기법이 적용된 모델 간의 정량적 성능 변화를 분석하는 심도 있는 Ablation Study를 통해 각 모듈의 실질적인 기여도를 규명할 계획이다.

참고문헌

[1] A. Shewale, A. Mokhade, N. Funde, and N. D. Bokde, “An Overview of Demand Response in Smart Grid and Optimization Techniques for Efficient Residential Appliance Scheduling Problem,” *Energies*, Vol. 13, No. 16, 4266, 2020. https://doi.org/10.3390/en13164266

[2] J. R. Vazquez-Canteli and Z. Nagy, “Reinforcement Learning for Demand Response: A Review of Algorithms and Modeling Techniques,” *Applied Energy*, Vol. 235, pp. 1072-1089, 2019. https://doi.org/10.1016/j.apenergy.2018.11.002

[3] K. Kwon and H.-B. Lee, “Performance Analysis of a Hybrid MSTL-SARIMAX Model for Multiple Seasonality Power Load Forecasting,” *Journal of Digital Contents Society*, Vol. 26, No. 12, pp. 3497-3505, December 2025. https://doi.org/10.9728/dcs.2025.26.12.3497

[4] F. Ruelens, B. J. Claessens, S. Vandael, B. De Schutter, R. Babuška, and R. Belmans, “Residential Demand Response of Thermostatically Controlled Loads Using Batch Reinforcement Learning,” *IEEE Transactions on Smart Grid*, Vol. 8, No. 5, pp. 2149-2159, September 2017. https://doi.org/10.1109/TSG.2016.2517211

[5] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, ... and D. Wierstra, “Continuous Control with Deep Reinforcement Learning,” in *Proceedings of the Internatinal Conference Learning Representations (ICLR)*, 2016. https://doi.org/10.48550/arXiv.1509.02971

[6] J. Yoon, D. Jarrett, and M. van der Schaar, “Time-Series Generative Adversarial Networks,” in *Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS)*, Vancouver, Canada, pp. 5508-5518, 2019.

[7] J. Ho, A. Jain, and P. Abbeel, “Denoising Diffusion Probabilistic Models,” in *Proceedings of the 34th Conference on Neural Information Processing Systems (NeurIPS)*, Vancouver, Canada, pp. 6840-6851, 2020. https://doi.org/10.48550/arXiv.2006.11239

[8] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, 5th ed. Hoboken, NJ: John Wiley & Sons, 2015. https://doi.org/10.1002/9781118619193

[9] S. Hochreiter and J. Schmidhuber, “Long Short-Term

Memory," *Neural Computation*, Vol. 9, No. 8, pp. 1735-1780, 1997. https://doi.org/10.1162/neco.1997.9.8.1735

[10] H. Wu, J. Xu, J. Wang, and M. Long, "Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting," in *Proceedings of the 35th Conference on Neural Information Processing Systems (NeurIPS)*, pp. 22419-22430, 2021. https://doi.org/10.48550/arXiv.2106.13008

[11] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, ... and D. Hassabis, "Human-Level Control Through Deep Reinforcement Learning," *Nature*, Vol. 518, 7540, pp. 529-533, 2015. https://doi.org/10.1038/nature14236

[12] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal Policy Optimization Algorithms," arXiv:1707.06347, 2017. https://doi.org/10.48550/arXiv.1707.06347

[13] Y. Tashiro, J. Song, Y. Song, and S. Ermon, "CSDI: Conditional Score-Based Diffusion Models for Probabilistic Time Series Imputation," in *Proceedings of the 35th Conference on Neural Information Processing Systems (NeurIPS)*, pp. 24804-24816, 2021. https://doi.org/10.48550/arXiv.2107.03502

권기현(Kihyeon Kwon)

1993년 : 강원대학교 컴퓨터과학과(학사)

1995년 : 강원대학교 대학원 컴퓨터과학과(석사)

2000년 : 강원대학교 대학원 컴퓨터과학과(박사)

2002년～현 재: 강원대학교 교수

※관심분야 : AIoT, 에너지 데이터 분석

이형봉(Hyung-Bong Lee)

1984년 : 서울대학교 계산통계학과(학사)

1986년 : 서울대학교 대학원 계산통계학과(석사)

2000년 : 강원대학교 대학원 컴퓨터과학과(박사)

1986년～1994년: LG전자 컴퓨터연구소

1994년～1999년: 한국디지털(주)

2004년～2025년: 강릉원주대학교 교수

2026년～현 재: 강원대학교 교수

※관심분야 : 무선 통신 (Wireless Networks), 센서 네트워크 (Sensor Networks), 임베디드 시스템 (Embedded Systems), 사물 인터넷 (IoT)