스마트 그리드 환경에서 비용 절감을 위한 강화학습 기법 성능 비교

노하진[1], 임유진[2]
[1]숙명여자대학교 IT공학과 석사과정
[2]숙명여자대학교 인공지능공학부 교수
hajins@sookmyung.ac.kr, yujin91@sookmyung.ac.kr

Performance Comparison of Reinforcement Learning for Cost Savings in Smart Grid

Hajin Noh[1], Yujin Lim[2]
[1]Dept. of IT Engineering, Sookmyung Women's University
[2]Div. of Artificial Intelligence Engineering, Sookmyung Women's University

요 약

IT 기술이 발전하며 실시간 전력 수요량 및 가격 등을 파악할 수 있는 스마트 그리드가 주목을 받고 있다. 스마트 그리드 환경에서는 에너지 저장 장치를 이용하여 소비자의 경제적 부담을 덜어낼 뿐만 아니라 에너지를 효율적으로 사용할 수 있다. 본 연구에서는 이러한 목표를 위해 과거 2시간 동안의 부하량 및 가격을 바탕으로 에너지 저장 장치의 충전 및 방전량을 결정하는 강화학습 알고리즘을 제안한다. 또한, 여러 강화학습 기법의 성능을 비교 분석한다.

1. 서론

오늘날, IT 기술이 발전함에 따라 스마트 그리드(Smart Grid)라는 개념이 등장하였다. 스마트 그리드란 지능형 기술을 통해 실시간 전력 소비량 및 공급량 등을 파악하고, 이를 통해 전력 생산 및 소비가 효율적으로 이루어지도록 하는 전력망을 말한다. 스마트 그리드 환경에서 에너지 저장 장치(Energy Storage System, ESS)는 전기를 저장하였다가 필요할 때 사용할 수 있게 함으로써 실시간으로 변동되는 전기 요금이나 재생 에너지의 불안정한 전력 공급에 유연하게 대응할 수 있도록 한다. 사용자는 ESS를 통해 전력 낭비를 줄일 수 있으며, 실시간 전력 가격이 저렴할 때 전기를 구매하고 가격이 높아졌을 때 판매 및 소비하여 비용을 절약할 수 있다.

2. 관련 연구

스마트 그리드에서 강화학습 기법을 이용하여 ESS의 에너지 관리를 최적화하는 연구로는 [1]이 있다. 해당 연구에서는 실시간 전력 요금이 변화하는 스마트 그리드 환경에서 Q-learning과 SARSA 기법을 이용하여 ESS가 최적의 시간에 충전 및 방전하도록 하는 알고리즘을 제안하였다. 이러한 알고리즘을 통해 소비자가 부담하는 비용을 크게 절감하는 결과를 보였다. [2]에서도 실시간 가격 변동 환경에서 DQN 기법을 이용하여 전기 요금이 저가일 때 매수하고 고가일 때 매도하도록 배터리 충전 및 방전을 결정하는 알고리즘을 제안하였다. 하지만, 이러한 연구에서 에이전트는 충전, 유지 또는 방전 여부만 결정하며, 그 전력량은 항상 일정하다는 한계가 있다. [3]에서는 Q-learning 기법을 통해 열에너지 저장 시스템과 배터리 저장 시스템을 사용하는 환경에서 ESS의 충·방전량을 최적화하여 최대 부하 시간대의 부하량을 감소시켰다. 하지만, 기준에 따라 일정한 가격이 부과되는 환경으로, 가격이 실시간으로 변동되는 환경에 적용하는 데에는 한계가 있다.

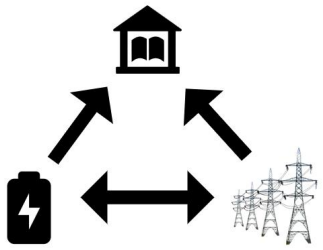
본 연구에서는 강화학습 에이전트가 충전 및 방전 행위를 결정할 뿐 아니라 얼마나 충·방전할 것인가를 선택하도록 한다. 또한, ESS의 최적 용량을 유지하여 배터리의 수명 연장을 목표로 하며, 동시에 소비자의 비용 부담을 줄일 수 있는 알고리즘을 제안한다.

3. 시스템 모델

3-1. 목표 및 환경 설정

에이전트의 목표는 ESS의 SoC(State Of Charge)를 일정 범위 내에서 유지하며 실시간 전력 가격을

고려하여 누적 전기 요금을 절약하는 것이다.



(그림 1) 스마트 그리드 구조

실험 환경인 스마트 그리드 구조는 (그림 1)과 같다. 건물은 ESS로부터 전력을 공급받을 수 있고, 그리드로부터 전력을 구매할 수 있다. ESS는 그리드로부터 전력을 구매 또는 판매할 수 있다. 그리드 역시 ESS로부터 전력을 구매하거나 판매할 수 있다.

실시간 부하량은 싱가포르 국립 대학교 건물에서 수집한 ROBOD[4] 데이터 중, 도서관에서 발생한 조명, 플러그, HVAC(Heating, Ventilation and Air Conditioning) 부하 데이터를 사용하였다. 해당 데이터는 5분 간격으로 수집되었으며, 2021년 9월부터 12월까지 주말 및 결측일을 제외한 47일간의 데이터로 구성되어있다. 실시간 가격은 EMC(Energy Market Company)[5]에서 30분 단위로 제공하는 싱가포르 국립 전력 시장(NEMS)의 실시간 가격 데이터를 5분 단위로 증강하여 사용하였다.

<표 1> ESS 환경 설정값

표기	값(kWh)
ESS_{cap}	6.8
ESS_{init}	2.72
$ESS_{\min}$	1.36
$ESS_{\max}$	5.44

<표 1>은 ESS 환경의 설정값이다. ESS의 총용량 ESS_{cap}는 시간대별 평균 부하량 중 가장 큰 값을 반올림한 값으로 설정하였다. ESS의 초기 용량 ESS_{init}은 ESS_{cap}의 40% 값이다. 또한, 비상 전력 보장 및 배터리의 수명 연장 등 효율적인 사용을 위해 ESS가 유지해야 할 목표 범위는 ESS의 20~80%으로, 각각 $ESS_{\min}$, $ESS_{\max}$와 같다.

3-2. 제안하는 알고리즘

실험에서 사용한 강화학습 기법은 DDPG(Deep Deterministic Policy Gradient), TD3(Twin Delayed Deep Deterministic Policy Gradient), SAC(Soft Actor-Critic)이다. DDPG는 이산 행동 공간 기반 알고리즘인 DQN(Deep Q-Networks)을 확장한 것으로, 연속 행동 공간에 적용할 수 있다. TD3는 DDPG를 개선한 알고리즘으로, 두 개의 Q-함수를 사용하여 DDPG보다 안정적으로 학습할 수 있다. SAC는 엔트로피를 추가하여 무작위성을 높인 알고리즘으로, 안정적인 정책을 학습할 수 있다. 데이터는 5분 단위로 구성되었으며, 24개의 데이터, 즉 2시간 동안의 데이터를 하나의 에피소드로 간주하여 학습하였다.

상태 공간은 현재 에피소드인 ep, 현재 시간 t, 과거 2시간 동안의 부하량 집합 $\{L_{t-24}, ..., L_{t-1}\}$, 현재 ESS의 충전 상태인 SoC_t, 과거 2시간 동안의 실시간 가격 집합 $\{P_{t-24}, ..., P_{t-1}\}$으로 이루어져 있으며, [식 1]과 같다.

$$S = [ep, t, \{L_{t-24}, ..., L_{t-1}\}, SoC_t, \{P_{t-24}, ..., P_{t-1}\}] \quad (1)$$

행동 공간은 [식 2]로, 에이전트는 [-0.15, 0.15]의 범위에서 충·방전량 a를 결정하도록 하여 한 번에 너무 많은 양을 충·방전하지 않도록 한다. 양수는 충전을 의미하며, 음수는 방전을 의미한다. 0은 해당 상태를 유지하는 것을 의미한다. 실제 충전 및 방전되는 전기량을 의미하는 $action$은 [식 3]과 같다.

$$a = [-0.15, 0.15] \quad (2)$$

$$action = ESS_{cap} \times a \quad (3)$$

새로운 SoC를 계산하는 식은 [식 4]와 같다. 단, 에이전트는 SoC_t를 초과하여 방전할 수 없고, ESS_{cap}을 초과하여 충전할 수 없다.

$$SoC_{t+1} = SoC_t + action \quad (4)$$

보상 식은 $penalty1$, $penalty2$, $penalty3$, cost 네 가지 구성 요소로 이루어져 있다. 먼저, $penalty1$은 [식 5]와 같으며, 남은 전력량보다 많이 방전하는 경우 잘못된 행동임을 알리기 위해 초과한 방전량만큼을 부과한다. $penalty1$의 범위는 (0, 1.02]이다.

$$penalty1 = |action| - SoC_t \quad (5)$$

$penalty2$는 [식 6]과 같으며, 목표 범위를 벗어난 경우, 벗어난 만큼을 부과한다. $penalty2'$는 $penalty2$를 (0, 1] 범위로 정규화한 값이다.

$$penalty2 = \begin{cases} SoC_{\min} - SoC_{t+1} \ (\text{if } SoC_{t+1} < SoC_{\min}) \\ SoC_{t+1} - SoC_{\max} \ (\text{if } SoC_{t+1} > SoC_{\max}) \end{cases} \quad (6)$$

$penalty3$은 [식 7]과 같이 $penalty2'$의 값을 부과하며, 목표 범위를 벗어났을 때, 에이전트가 목표 범

위 내로 복귀하도록 하는 역할을 한다. 범위 역시 $penalty2'$와 같다.

$$penalty3 = penalty2' \quad (7)$$

cost는 [식 8]과 같고, 이익일 때는 양수, 손해일 때는 음수값을 갖는다. $action'$는 $action$을 [0, 1]로 정규화한 값이고, 현재 가격 P_t의 범위는 [0.10155. 2.03551]이다.

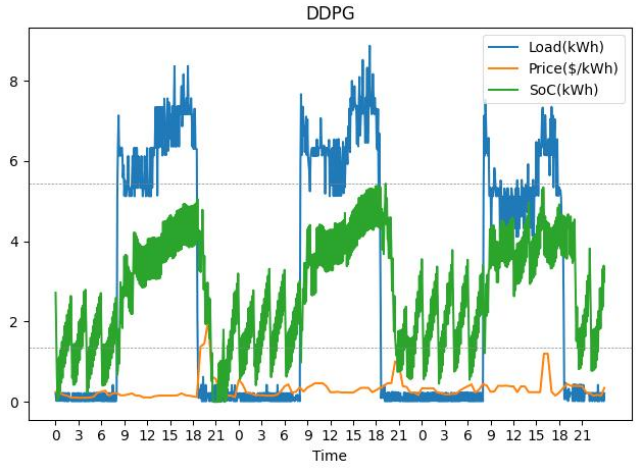
$$\text{cost} = action' \times P_t \quad (8)$$

모든 요소를 고려한 최종 식은 [식 9]과 같으며 (-2, 1) 범위를 갖는다.

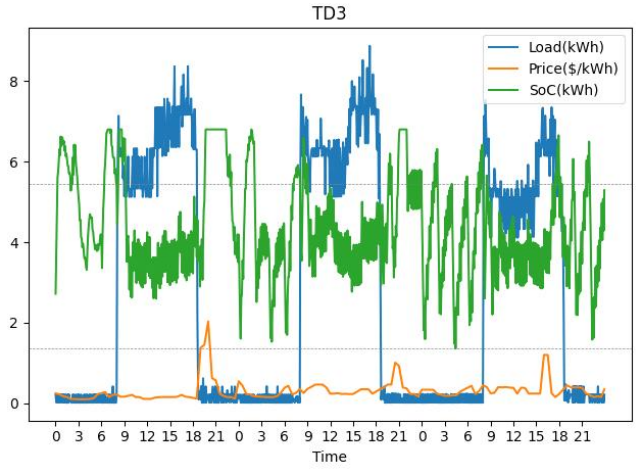
$$reward = \frac{(-penalty1 - penalty2' - penalty3 + \text{cost})}{3} \quad (9)$$

4. 실험 결과

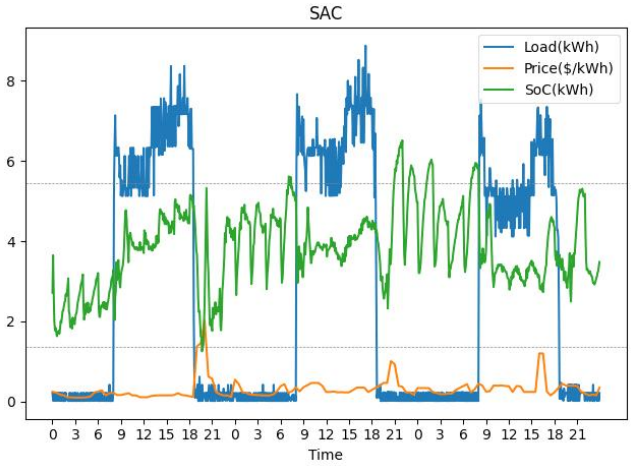
다음 (그림 1), (그림 2), (그림 3)은 동일한 테스트 데이터인 2021년 12월 21일부터 23일까지에 대한 DDPG, TD3, SAC 동작 그래프이다. 점선은 $ESS_{\min}$와 $ESS_{\max}$를 나타낸다.



(그림 1) DDPG 실험 결과



(그림 2) TD3 실험 결과



(그림 3) SAC 실험 결과

DDPG는 부하가 적은 심야 시간에 충·방전을 반복하며 부하가 많은 낮 시간대에는 대체로 충전하는 것을 볼 수 있다. TD3는 DDPG보다 변화하는 폭이 크며, DDPG의 SoC는 심야 시간대에 $ESS_{\min}$ 전후에서 유지하는 반면 TD3는 항상 $ESS_{\min}$ 이상을 유지한다. SAC의 SoC 역시 심야 시간에 거의 $ESS_{\min}$ 이상을 유지하나, 그 폭이 TD3에 비해 작다. 또한, TD3를 제외한 DDPG와 SAC에서는 다른 시간에 비해 가격이 크게 상승하는 시간에 방전되는 것을 뚜렷하게 관찰할 수 있다.

<표 2> 기법별 3일간 누적 전기 요금

기법	누적 요금($)		
	1일차	2일차	3일차
DDPG	138.37	413.98	703.38
TD3	130.87	402.44	690.38
SAC	132.70	405.54	694.14

<표 2>는 기법별 3일간 누적 전기 요금이다. TD3 기법을 적용하였을 때, 690.38($)로 가장 적은 금액을 기록하였으며, SAC 기법이 694.14($), DDPG 기법이 703.38($)를 기록하여 TD3가 가장 경제적임을 알 수 있다.

<표 3> 기법별 3일간 목표 범위 이탈 누적 횟수

기법	목표 범위 이탈 횟수		
	1일차	2일차	3일차
DDPG	45	122	157
TD3	70	161	191
SAC	2	28	50

<표 3>은 기법별 3일간 목표 범위를 벗어난 누적 횟수이다. TD3가 이탈 횟수가 가장 잦았고, DDPG, SAC가 그 뒤를 이었다. 하지만, DDPG는 대체로 $ESS_{\min}$보다 작은 값으로 이탈하였고, TD3와 SAC는 $ESS_{\max}$보다 큰 값으로 이탈하였다는 점에서 차이가

있다. 따라서, 심야 시간에 비상 전력이 필요할 경우, DDPG보다 TD3나 SAC 기법이 더 적합함을 알 수 있다.

5. 결론

본 연구에서는 스마트 그리드 환경에서 ESS의 효율적이고 경제적인 운영을 위해 ESS의 SoC와 실시간으로 변화하는 전력 가격을 고려하여 충·방전량을 조절하는 강화학습 알고리즘을 제안하였다. 또한, 동일 MDP를 DDPG, TD3, SAC 기법에 적용하여 다양한 에이전트의 동작을 비교하였다. 실험 결과, 대체로 SoC를 목표 범위 내에서 유지하며, 실시간 가격이 높을 때 충전하고 저렴할 때 방전하는 패턴을 확인할 수 있었다. 비용 절감 면에서는 TD3가 가장 뛰어난 성능을 보였으며, SAC, DDPG가 그 뒤를 따랐다. 하지만, 목표 범위 이탈한 횟수는 SAC가 현저히 적은 횟수를 기록하였고, DDPG와 TD3 순으로 목표 범위를 잘 유지하는 것으로 나타났다.

이러한 연구 결과를 통해 소비자들은 스마트 그리드 환경에서 더욱 합리적인 소비를 할 수 있을 것으로 보이며, 더 나아가 에너지 낭비도 줄일 수 있을 것으로 기대된다. 향후 연구에서는 스마트 그리드에서 ESS와 함께 많이 쓰이는 태양광 에너지를 고려하여 전력 공급이 불안정한 동적인 상황에서 ESS의 SoC 최적화와 비용 저감을 목표로 하는 연구를 진행하고자 한다.

사사문구

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 ICT혁신인재4.0 사업의 연구결과로 수행되었음 (IITP-2024-RS-2022-00156299)

참고문헌

[1] P. Bijapur, P. Chakradhar, M. Rishab, S. Srinivas and B. S. Nagabhushana, "Reinforcement Learning for Energy Storage Optimization in the Smart Grid," 2020 IEEE International Conference on Power Electronics, Smart Grid and Renewable Energy (PESGRE2020), Cochin, India, pp. 1-6, 2020.

[2] E. Brock, L. Bruckstein, P. Connor, S. Nguyen, R. Kerestes and M. Abdelhakim, "An Application of Reinforcement Learning to Residential Energy Storage under Real-time Pricing," 2021 IEEE PES Innovative Smart Grid Technologies - Asia (ISGT Asia), Brisbane, Australia, pp. 1-5, 2021.

[3] Z. Rostmnezhad and L. Dessaint, "Power Management in Smart Buildings Using Reinforcement Learning," IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), Washington, DC, USA, pp. 1-5, 2023.

[4] Tekler, Z.D., Ono, E., Peng, Y. et al. "ROBOD, Room-level Occupancy and Building Operation Dataset," Building Simulation. vol. 15, no. 12, pp. 2127 - 2137, 2022.

[5] Energy Market Company, https://www.home.emcsg.com/