비용 절감을 위한 강화학습 기반 건물 내 ESS 충·방전 스케줄링 기법

임세헌, 손예지, 윤성국
숭실대

Reinforcement Learning-based ESS Scheduling for Cost Minimization

Se-Heon Lim, Ye-ji Son, Sung-Guk Yoon
Soongsil University

Abstract - 온실가스 감축을 위한 노력의 일환으로 소비자 단의 에너지 관리가 주목받고 있다. 특별히 현대 사회에서는 도시화로 인한 건물에서의 에너지 소비가 전체의 약 24%를 차지하고 있기에 건물의 효율적인 에너지 사용이 중요하다. 건물 내 에너지를 효율적으로 사용하기 위해 냉난방과 조명 부하 등을 실시간으로 조정하는 방식과 ESS(Energy Storage System)를 이용하는 방식 등이 고려되고 있다. 본 논문에서는 건물에서 에너지의 비용을 줄이기 위해 강화학습 기반의 ESS 충·방전 스케줄링 알고리즘을 제안한다. 사례연구에서 제안한 강화학습 기반 충·방전 스케줄링 기법이 선행연구에서의 방식보다 유리한 상황이 있다는 것을 보였다.

1. 서 론

지구온난화라는 전 세계적인 위협에 대응하기 위해 세계 각국에서는 온실가스 감축을 진행하고 있다. 이에 따라 효율적인 에너지 소비의 중요성이 대두되었으며, 정부는 에너지 성능을 평가하는 다양한 제도를 도입하여 건물 에너지 부문의 효율 향상을 도모하고 있다.

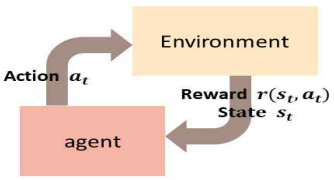
따라서 에너지 성능 향상을 위한 건물 에너지 관리에 관한 다양한 연구가 진행되었다. 특히 본 논문에서 참고한 선행연구[1]에서는 ESS(Energy Storage System), PV(Photovoltaic), EV(Electrical Vehicle)을 이용하여 스마트 빌딩 내 에너지관리시스템을 구현하고, 강화학습 알고리즘의 성능을 검증하였다.

본 논문에서는 선행연구를 토대로 하여 기존 강화학습 알고리즘의 보상함수를 개선하여 방전량은 감소하고 방전 빈도수는 늘리도록 유도하였다. 제안하는 방법을 통해 보다 안정적이고 경제적인 ESS 운용이 가능하다. 사례연구에서는 숭실대학교 전력량 데이터를 이용하여 선행연구의 방식과 본 논문에서 제안하는 방식의 이익률을 비교하고, 각 경우의 ESS 스케줄링의 차이를 그래프로 도시하였다.

2. 본 론

2.1 강화학습

강화학습은 기계학습의 한 영역으로 에이전트가 환경(Environment) 내에서 가능한 행동 중 누적 보상을 최대화하는 행동을 취하게끔 설계된 알고리즘이다. 에이전트가 알고리즘에 의해 행동(action)을 취하면 그에 따르는 보상(reward)을 받으며, 행동 후에는 에이전트가 처한 상황(state)이 바뀌게 된다. 이 같은 행위를 반복하면서 에이전트는 누적된 보상을 최대화 하는 정책을 학습한다.[2]



〈그림1〉 강화학습

2.2 에너지 관리 모델 구성

〈표1〉 명명법

$price_t$	가격	s_t	state	Δe	충전량
p_t	소비전력	a_t	action	ESS_t	ESS SOC
A	action set	$reward(a_t)$	보상함수	$ESS_{\max}$	ESS 저장용량

〈표2〉 여름철 전기 요금 (고압 A 선택II)

구분	시간	가격
경부하	23:00~09:00	45.3(원/kWh)
중부하	09:00~10:00 12:00~13:00 17:00~23:00	90(원/kWh)
최대부하	10:00~12:00 13:00~17:00	155.9(원/kWh)

$$s_t = [ESS_t, p_t, price_t] \in S \quad (1)$$

$$A = \{\text{구매}, \text{충전}, \text{방전}\} \quad (2)$$

$$reward(a_t) = \text{cost}(t) \quad (3)$$

에이전트의 상황(state)은 식(1)와 같이 ESS의 SOC(State Of Charge), 소비전력, 현재 전기요금[1]로 구성되어 세 가지 요소로서 현재 상황을 특정할 수 있다. 6월에 대해서만 사례연구를 진행하였기 때문에 여름철 전기요금표만 표기하였다. ESS 에이전트가 시간 t에서 취할 수 있는 행동(action)은 식(2)에서와 같이 구매, 충전, 방전 중 하나이다. 에이전트가 행동을 취하면 그에 따르는 보상을 받고 다음 상황(state)으로 이동하게 된다. 본 논문에서 사용한 보상은 행동을 취한 후 발생하게 되는 비용으로 식(3)과 같이 산정하였다.

다음에 소개되는 2.2.1에서는 본 논문에서 설정한 조건식들을 기존방식과 비교하여 설명한다.

2.2.1 제안방식

각 행동 시 ESS의 충방전량은 식(4)와 같다. 선행 연구에서는 짧은 피크부하 시간대를 고려하여 방전량을 소비전력량으로 하여 짧은 시간 내에 많은 방전량을 유도하였다. 이를 심화하여 본 논문에서는 ESS 의 안정적인 운용을 목적으로 하여 보다 적은 방전량을 유도하기 위해 ESS의 충전량 함수를 식(4)처럼 변경하였다. 식(4)의 조건에 의해 시간 t 에 가능한 행동도 제한되는데 식(5)에서 action space를 확인할 수 있다.

$$C(a_t) = \begin{cases} 0 & \text{if } a_t = \text{구매} \\ \Delta e & \text{if } a_t = \text{충전} \\ -\min(p_t, ESS_t) & \text{if } a_t = \text{방전} \end{cases} \quad (4)$$

1) $price_t$로 본 논문에서는 <표2>과 같이 한전 교육용 TOU(Time of Price) 요금제를 따른다.

$$A_{s_t}=\begin{cases}\text{구매,충전} & \text{if } ESS_t=0\\ \text{구매,충전,방전} & \text{if } 0<ESS_t\le ESS_{\max}-\Delta e\\ \text{구매,방전} & \text{if } ESS_{\max}-\Delta e<ESS_t\le ESS_{\max}\end{cases} \quad (5)$$

$$r(s_t,a_t)=\begin{cases}p_t\cdot price_t & \text{if } action=\text{구매}\\ (p_t+\Delta e)\cdot price_t & \text{if } action=\text{충전}\\ (p_t-\min(p_t,ESS_t))\cdot price_t & \text{if } action=\text{방전}\end{cases} \quad (6)$$

에이전트가 행동(action)을 취하면 그에 따르는 보상을 받게 되는데 본 논문에서 설정한 보상은 행동 시 발생하는 비용으로 산정되었다. 보상을 비용으로 줬기 때문에 에이전트는 행동을 취할 때 누적 비용의 최소화를 위해 Q값 중 가장 작은 값에 해당하는 행동을 선택한다.

2.3 Q-learning

Q learning 이란 강화학습 기법 중 하나로, Q learning은 현재 상태에서 수행하는 행동에 대한 가치함수인 행동 가치 함수를 학습하는 것을 말한다. Q-learning 의 학습이 종료되면 에이전트는 각 상태에서 최적의 Q 값에 해당하는 행동을 수행하게 된다.[3] 본 논문에서 사용한 행동가치 함수는 식(8)에 의해 업데이트 된다.

$$Q(s_t,a_t)\leftarrow(1-\alpha)Q(s_t,a_t)+\alpha[r(s_t,a_t)+\gamma\cdot\min_{a\in A_{s_{t+\Delta t}}}Q(s_{t+\Delta t},a)] \quad (8)$$

2.3.1 $\epsilon-greedy$ 기법

$\epsilon-greedy$ 기법은 Q-learning 알고리즘에 사용한 기법으로 상황(state)에서 에이전트가 행동을 취할 때 $1-\epsilon$ 확률로 지금까지 업데이트된 Q값의 최소값(혹은 최대값)에 해당하는 행동을 하거나 ϵ 확률로 무작위로 행동하여 다음 Q 값을 업데이트 하는 방식을 말한다. 이는 현재 최적의 행동이 정확한 답이 아닐 수 있기 때문에, 더 좋은 해결책을 찾기 위해 다른 행동들을 무작위로 시도해보는 기법이다.[4]

2.4 결과

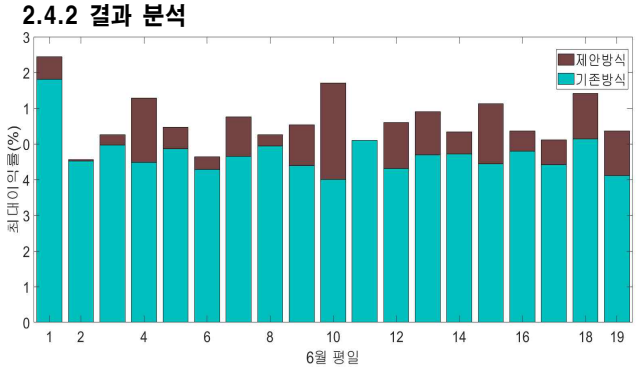
2.4.1 데이터 정보

입력 데이터로서 숭실대학교 6월평일 전력량 데이터(19일)를 사용하여 학습을 진행하였다. 학습에 쓰이는 주요 입력 변수들은 <표3>과 같다.

<표 3> 입력 변수

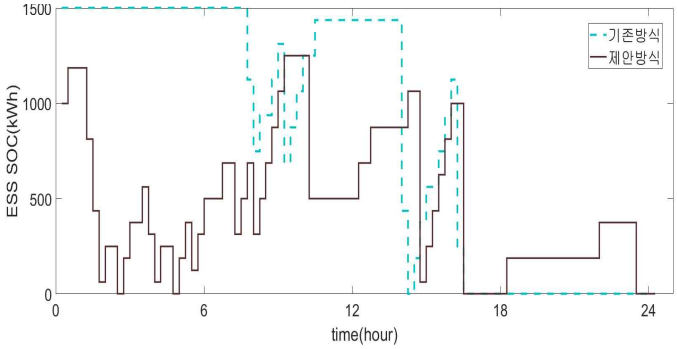
이름	값	이름	값
ESS 충전량	$\Delta e=187.5kWh$	학습 속도	0.1
ESS 저장용량	$ESS_{\max}=1500kWh$	할인율	0.95
ESS SOC 초기값	$ESS_0=1000kWh$	ϵ	0.2
시간 주기	$\Delta t=15\text{min}$	반복 횟수	10000

2.4.2 결과 분석



<그림 3> 최대 이익률 비교

제안하는 기법의 성능을 분석하기 위해 이익률을 ESS를 사용하지 않았을 때와 비교하여 비용이 감소되는 비율로 정의한다. 매 반복 마다 이익률을 계산하여 그 중 최대값을 최대이익률로 산정하였다. <그림 3>은 평일 19일에 대하여 제안방식과 기존방식의 최대이익률을 비교해 놓은 그래프이다. 제안방식은 19일 중 18일에서 기존방식보다 높은 이익률을 보였고 제안방식과 기존방식의 최대 이익률 평균은 각각 약2.82%, 2.36%로 제안방식이 더 높은 이익률을 보였다.<그림 4>는 평일 중 임의의 날(6월 24일)의 제안방식의 ESS SOC 스케줄링을 보여준다.



<그림 4>ESS SOC Scheduling 예시

그래프에서 확인할 수 있듯이, 제안방식에서는 시간 t 에서의 방전량이 기존방식보다 줄어든 것을 확인 할 수 있다. 한 번에 과다한량의 전기를 방전하게 되면 ESS에 손상을 줄 수 있으므로 본 논문에서 제안하는 방식은 ESS의 안정적인 운용을 기할 수 있다는 점에서 유리하다.

3. 결 론

본 논문에서는 숭실대학교 전력량 데이터를 이용하여 강화학습 기반의 ESS 충·방전 스케줄링을 수행하였다. 선행연구에서 적용한 방식에서 방전량을 줄여보고자 함수를 변형하여 연구를 진행하였다. 결과분석에서 제안하는 방식이 기존방식에 비해 높은 이익률을 보이는 것을 확인하였다. 추가적으로 기존방식은 한 번에 과다한량의 전기를 방전하는 반면, 제안방식은 순간 방전량이 낮기에 제안하는 방식을 통해 ESS를 안정적으로 동작할 수 있다.

후속 연구로서 현실성을 반영해, 제한된 PCS용량을 고려하여 PCS의 용량이상을 방전하지 못하도록 행동 집합을 설정하는 것을 고려할 수 있다.

[참 고 문 헌]

[1] Kim, Sunyong, and Hyuk Lim. "Reinforcement learning based energy management algorithm for smart energy buildings." *Energies*, 2018.11.8

[2] Otterlo, Martijn van and Marco Wiering, "*Reinforcement Learning and Markov Decision Processes*". Springer, 3-42, 2012.

[3] Watkins, Christopher JCH, and Peter Dayan. "Q-learning." *Machine learning*, 279-292, 1992.

[4]TOKIC, Michel; PALM, Günther."Value-difference based exploration: adaptive control between epsilon-greedy and softmax," *Proc. Annual Conference on Artificial Intelligence*, 335-346,2011.

감사의 글

본 논문은 한국전력공사의 2018년 착수 에너지 거점대학 클러스터 사업의 지원을 받아 수행된 연구임.(Grant number. R18XA04)