

https://doi.org/10.7236/JIIBC.2021.21.5.215
JIIBC 2021-5-29

스마트 빌딩 시스템을 위한 심층 강화학습 기반 양방향 전력거래 협상 기법

Bi-directional Electricity Negotiation Scheme based on Deep Reinforcement Learning Algorithm in Smart Building Systems

이동구*, 이지영*, 경찬욱*, 김진영**

Donggu Lee*, Jiyoung Lee*, Chanuk Kyeong*, Jin-Young Kim**

요약 본 논문에서는 스마트 빌딩 시스템과 전력망이 각각의 전력거래 희망가격을 제안하고 조정하는 양방향 전력거래 협상 기법에 심층 강화학습 기법을 적용한 전력거래 기법을 제안한다. 심층 강화학습 기법 중 하나인 deep Q network 알고리즘을 적용하여 스마트 빌딩과 전력망의 거래 희망가격을 조정하도록 하였다. 제안하는 심층 강화학습 기반 양방향 전력거래 협상 알고리즘은 학습과정에서 평균 43.78회의 협상을 통해 가격 협의에 이르는 것을 실험을 통해 확인하였다. 또한, 본 연구에서 설정한 협상 시나리오에 따라 스마트 빌딩과 전력망이 거래 희망가격을 조정하는 과정을 실험을 통해 확인하였다.

Abstract In this paper, we propose a deep reinforcement learning algorithm-based bi-directional electricity negotiation scheme that adjusts and propose the price they want to exchange for negotiation over smart building and utility grid. By employing a deep Q network algorithm, which is a kind of deep reinforcement learning algorithm, the proposed scheme adjusts the price proposal of smart building and utility grid. From the simulation results, it can be verified that consensus on electricity price negotiation requires average of 43.78 negotiation process. The negotiation process under simulation settings and scenario can also be confirmed through the simulation results.

Key Words : Electricity Negotiation, Deep Learning, Deep Reinforcement Learning, Smart Building Systems,

1. 서론

세계적으로 전력 거래 시장은 기존 독점적 시장 형태에서 다양한 사업체가 경쟁하는 개방된 형태로 변화하고 있다^[1]. 이에 따라 전력 거래 경매와 다자간의 전력거래

등 다양한 방법이 연구되고 있다^{[2], [3]}.

최근 다양한 기술 분야에 인공지능 기법을 도입한 연구가 활발히 진행되고 있다. 인공지능 기법은 크게 지도 학습, 비지도 학습, 강화학습으로 구분된다^[4]. 강화학습은 주어진 task를 수행하는 agent가 reward를 최대화

*준회원, 광주대학교 전자융합공학과

**정회원, 광주대학교 전자융합공학과

접수일자 2021년 5월 5일, 수정완료 2021년 9월 1일

게재확정일자 2021년 10월 8일

Received: 5 May, 2021 / Revised: 1 September, 2021 /

Accepted: 8 October, 2021

**Corresponding Author: Jin Young Kim, jinyoung@kw.ac.kr

Dept. of Electronic Convergence Engineering, Kwangwoon Univ., Korea

하도록 task를 학습시키는 기법으로, 각종 최적화 문제를 task로 설정하여 성능을 최적화할 수 있다^[5]. 특히 심층 신경망을 강화학습 기법에 적용한 심층 강화학습은 주어진 환경에서 최적의 선택을 자동화하도록 설계할 수 있다.

[6]의 연구에서는 심층 강화학습 알고리즘 중 하나인 deep Q network(DQN) 알고리즘을 제안하였다. DQN 알고리즘은 심층 강화학습의 발산 문제를 해결하고 수렴성을 개선하였다. 해당 연구에서는 convolutional neural network(CNN)^[7]을 적용하여 Atari 비디오 게임 화면을 인식하고 게임 스코어를 최대화하도록 agent를 학습시켰다. Agent가 CNN 기반 영상인식 모델을 통해 게임 환경을 인식하여 최적의 플레이를 위한 컨트롤 신호를 발생하도록 하여 사람 수준의 게임 스코어를 기록했다고 한다.

[8]의 연구에서는 particle swarm optimization(PSO) 알고리즘 기반 전력거래 최적화 기법을 제안하였다. 해당 연구에서는 스마트 빌딩과 전력망이 거래하는 상황에서 수치해석 기법을 통해 판매자와 구매자의 거래 희망가격을 조정하도록 하였다. 정해진 시간 동안 구매자와 판매자는 거래 희망가격을 제안하고 역제안하며 조정하고, 거래 희망가격이 같아질 경우 거래가 체결된다.

전력 거래 협상에서 협의에 보다 빠르게 도달하기 위해서는 협상 주체들이 서로의 상호 기댓값의 차이를 줄이는 것이 관건이라고 알려져 있다^{[9], [10]}. 본 논문에서는 [8]의 연구와 같이 스마트 빌딩과 전력망이 거래 희망가격을 주고받으며 가격 협의에 이르는 협상 모델에 심층 강화학습 알고리즘인 DQN 알고리즘을 적용하여 가격 협의에 이르는 기간을 단축하고자 하였다.

본 논문의 구성은 다음과 같다. 이어지는 2장에서는 스마트 빌딩 시스템과 심층 강화학습을 적용한 양방향 전력거래 협상 기법을 기술한다. 3장에서는 실험 환경과 실험에 사용한 파라미터를 설명하고 이에 따른 실험결과를 제시하고 분석한다. 마지막으로 4장에서는 실험결과를 요약하고 논문을 마무리하며 결론을 제시한다.

II. 시스템 모델

1. 스마트 빌딩 시스템 모델

스마트 빌딩 시스템은 최첨단 빌딩 관리 시스템으로써 다양한 정보처리 기술과 통신기술을 활용하여 빌딩 내의 환경을 자동적으로 최적화하는 특징이 있다^[11]. 또한 스

마트 빌딩 시스템은 보조 동력원으로써 태양광 패널과 윈드 터빈을 이용할 수 있고, 이를 통해 발생된 보조 동력을 저장하는 배터리가 탑재될 수 있다. 전력 사용량을 절감하기 위해 스마트 빌딩은 친환경 동력원인 보조 동력원을 우선적으로 사용하고, 이것이 부족하게 될 경우 전력망에서 전력을 충당하여 사용하게 된다. 반대로, 스마트 빌딩이 친환경 동력원으로부터 전력을 공급받고 남은 전력은 인근 스마트 빌딩 혹은 전력망으로 판매할 수 있다. 그림 1은 스마트 빌딩 시스템 구조를 나타낸다.

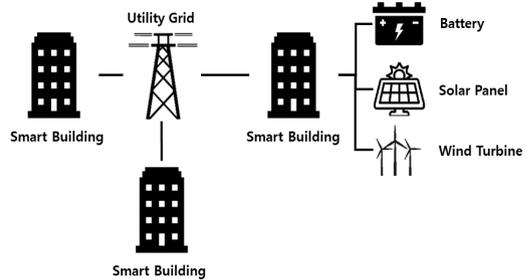


그림 1. 스마트 빌딩 시스템 구조.
Fig. 1. Smart building systems architecture.

스마트 빌딩은 [11]의 연구에서 제시된 실내 사용자의 comfort 레벨을 높은 수준에서 유지하도록 한다. 실내 사용자 comfort 레벨은 냉난방, 전등, 공기질 개선 시스템 등 스마트 빌딩의 환경에 관련된 요소이다. 본 논문에서는 스마트 빌딩은 임의의 실내 사용자의 comfort 레벨 기준을 항상 만족하도록 전력을 확보하고 유지하도록 하는 것을 가정하였다. 또한, 스마트 빌딩을 구매자, 전력망을 판매자로 가정하여 연구를 진행하였다.

2. 심층 강화학습 모델

심층 강화학습 기법에는 state, action, reward의 세 가지 요소가 정의되어야 한다. 본 논문에서 설정한 심층 강화학습의 세 가지 요소는 다음과 같다.

- $S = [s_1, s_2]$ 로 이루어져 있으며, s_1 은 스마트 빌딩의 거래 희망가격을 나타내며 s_2 는 전력망의 거래 희망가격을 나타낸다.
- $A = [a_1, a_2, \dots, a_{12}]$ 로 나타내며, 각 6개의 action은 s_1 과 s_2 를 주어진 step size 만큼 조정하도록 설정하였다. 예를 들면, a_1 은 s_1 을 step size 만큼 감소시키고, a_2 는 s_1 을 step size 만큼 증가시킨

다. 또한 a_3, a_4 는 a_1, a_2 에 비해 step size가 두 배가 되도록 설정하였고 a_5, a_6 은 세 배가 되도록 설정하였다. 위의 설정방식은 a_7 에서 a_{12} 까지 동일하게 적용하도록 설정하였다.

■ $R = 100 / |s_1 - s_2|$ 로 설정하여 s_1 과 s_2 의 차이가 작아질수록 reward가 커지도록 설정하였다. 이를 통해 스마트 빌딩과 전력망 사이의 거래 희망가격 협상을 가속화하도록 하였다.

3. 심층 강화학습 기반 양방향 전력거래 협상 기법

강화학습 agent는 Q 함수^[5]를 이용하여 누적 reward를 최대화 하는 방향으로 학습한다. 이때 Q 함수는 다음과 같이 표현된다.

$$Q(s, a) = E \left[\sum_{t=0}^{\infty} d^t R_t | s_t = s, a_t = a \right], \quad (1)$$

이때 t 는 time step, d 는 discount factor, $E[\cdot]$ 는 기댓값을 나타낸다. Discount factor는 time step에 따라 reward를 감소시키게 된다. Q 함수를 이용한 Q 학습 알고리즘의 경우 Q 함수의 값을 다음과 같은 수식을 통해 업데이트 한다.

$$Q(s, a) \leftarrow Q(s, a) + l [r_t + d \max_{a'} Q(s', a') - Q(s, a)] \quad (2)$$

이때 l 은 학습률, s' 과 a' 는 다음 time step의 state와 action을 의미한다. Q 학습 기법은 위의 수식을 토대로 Q 함수를 업데이트하며 최적화한다.

하지만 Q 학습 기법은 발산 및 수렴 불안정성의 문제가 있다고 알려져 있다. 이를 해결하기 위해 DQN 알고리즘이 [6]의 연구에서 처음 제시되었다. DQN 알고리즘은 1) 심층 신경망을 이용하고, 2) 메인 신경망과 타겟 신경망을 구분해서 운영하고, 3) replay memory를 사용한 학습 전략을 이용한다는 특징이 있다. 이를 통해 DQN 알고리즘은 아래의 오차함수를 통해 신경망을 학습시킨다.

$$L(\theta) = [Q(s, a; \theta) - \{r + d \max_{a'} \hat{Q}(s', a', \theta^-)\}]^2, \quad (3)$$

이때 θ 는 메인 신경망의 파라미터, θ^- 는 타겟 신경망의 파라미터, $\hat{Q}[\cdot]$ 는 타겟 신경망의 Q 함수값을 나타

낸다^[6].

전력 사용량이 물리는 피크 시간대에서 스마트 빌딩과 전력망의 희망 거래가격을 나타내는 s_1 과 s_2 는 각각 다음과 같이 설정된다^{[8], [12]}.

$$s_1 = C + \alpha(P_{hl} - C), \quad (4)$$

$$s_2 = D - \beta(D - P_{ll}), \quad (5)$$

이때 C 는 판매자의 reservation 가격이고, D 는 구매자의 reservation 가격, P_{hl} 은 시장에서의 가격 최대치, P_{ll} 은 시장에서의 가격 최저치를 의미하며, α 는 [0.85, 1] 영역에서의 임의의 소수, β 는 [0.5, 0.85] 범위에서의 임의의 소수를 나타낸다. 또한 action의 단위가 되는 step size는 다음과 같이 설정하였다.

$$|a_1| = \gamma^* EA_{s_1}, \quad (6)$$

$$|a_7| = \gamma^* EA_{s_2}, \quad (7)$$

이때 γ 는 step 상수, EA_{s_1} 과 EA_{s_2} 는 s_1 과 s_2 의 eagerness 값을 의미한다^[8].

심층 강화학습 agent는 주어진 round count 동안 state를 입력받아 최적의 action 선택지 출력을 낸다. 이때 exploration과 exploitation의 균형을 맞추기 위해 ϵ -greedy 알고리즘을 이용하였다. Action에 따라 state를 조정하고 s_1 과 s_2 의 차가 step 상수보다 작아질 때까지 조정을 반복한다. 이때 state값이 음수로 넘어가게 되는 경우 -100의 reward 패널티를 주고 해당 episode 내에서 reward 값이 최대가 되었던 state를 불러와 조정을 지속하도록 하였다. 조정 횟수 즉 round count가 모두 소진되어 episode가 끝나거나, s_1 과 s_2 의 차이가 step 상수보다 작아져 희망 거래가격 협의에 이르는 경우 episode가 종료되고 다음 episode를 시작하게 된다.

III. 실험 환경 및 결과

1. 실험 환경

본 연구의 실험은 메인 신경망과 타겟 신경망 모두 4계층 신경망을 이용하였으며 계층 사이에 dropout 레이어^[13]를 추가하였다. Discount factor d 는 0.9, 최대 episode와 episode당 조정 횟수(round count)는 모두

200으로 설정하였다. s_1 과 s_2 의 eagerness는 각각 0.1과 0.625, P_{hi} 은 11 cent/kWh, P_{li} 은 1 cent/kWh, C 와 D 는 2와 9로 설정하였다^[8]. Step 상수 γ 는 0.2로 설정하였으며, optimizer는 RMSProp optimizer를 사용하였다^[14].

2. 실험 결과

그림 2는 심층 강화학습 기반 양방향 전력거래 협상 기법의 episode에 따른 협상 체결 소요 round count를 나타낸다. 학습과정에서 초기 episode에서는 협의에 이르지 못하거나 160 이상의 협상 round가 진행된 이후에 협의에 이르는 경우가 나타난다. Episode가 진행될수록 협의에 이르는 데 필요한 round 수가 줄어들어 일정 수준에 수렴하는 경향을 보인다. 전체적으로 평균 43.78의 round가 필요한 것으로 실험결과를 통해 확인하였다.

그림 3은 그림 2에서 나타난 episode에 따른 협의에 필요한 round count 실험 결과 중에서 임의의 한 episode를 별도로 관찰한 결과를 보여준다. 각 episode 내에서 협의에 필요한 round count의 결과는 그림 2를 통해 알 수 있다. 그림 3의 예시의 경우 약 50번의 상호 희망 거래가격 제안과 역제안을 통해 거래가격 협의에 이르는 것을 볼 수 있다. 이때 전력망의 희망 거래가격 변동률보다 스마트 빌딩의 희망 거래가격 변동률이 높은 것으로 관측된다. 이는 주어진 피크 시간대에서 스마트 빌딩이 빌딩 내부의 comfort 레벨을 유지하기 위해 전력을 확보하려는 경향을 갖는 전력거래 협의 시나리오를 가정하였고, 이를 효과적으로 나타내기 위해 스마트 빌딩의 eagerness를 높게 설정하였기 때문이다. 다시 말해, 전력망보다 스마트 빌딩이 전력을 확보하려는 의향이 있고 이를 위해 희망거래 가격 측면에서 상대의 제안에 수용적으로 가격을 조정한다는 의미로 볼 수 있다.

IV. 결 론

본 논문에서는 심층 강화학습을 양방향 전력거래 모델에 적용하여 구매자와 판매자가 희망 거래가격을 조정하는 과정을 가속화하였다. 설정된 실험 환경 및 시나리오 내에서 심층 강화학습 모델은 평균 43.78회의 희망 거래 가격 협상 과정을 통해 협의에 이르는 것을 확인하였다. 추후 연구로는 협의에 필요한 협상 과정을 단축하기 위해 다양한 심층 강화학습 기법을 적용하는 것을 고려할 수 있다.

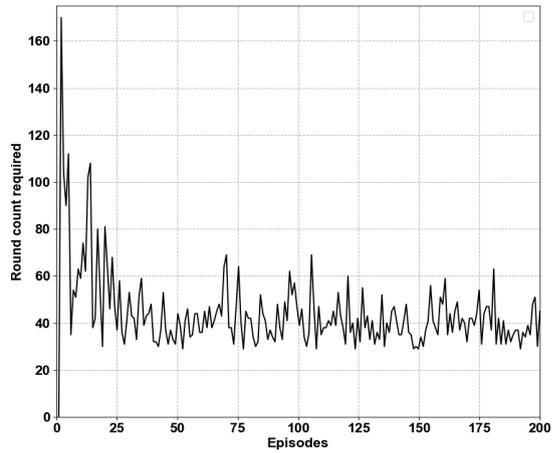


그림 2. Episode에 따른 협상 체결 round 수.
Fig. 2. Required round count over episodes.

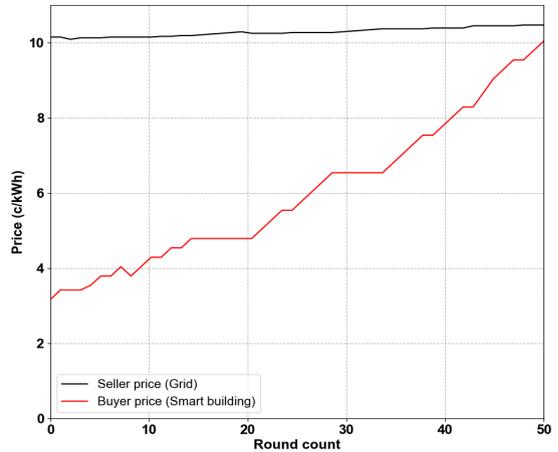


그림 3. Round count에 따른 협상 체결 과정.
Fig. 3. Negotiation process over round count.

References

- [1] S. Y. Al-Agtash and A. A. Al-Fahoum, "An evolutionary computation approach to electricity trade negotiation," *Advances in Engineering Software*, vol. 36, no. 3, pp. 173-179, March 2005.
- [2] S. Al-Agtash and R. Su, "Economic efficiency of coordinated multilateral trades in electricity markets," *International Journal of Electrical Power and Energy Systems*, vol. 24, no. 10, pp. 843-850, Dec. 2002.
- [3] Al-Agtash and R. Su, "Economic efficiency of pool coordinated electricity markets," *International Journal of Electrical Power and Energy Systems*, vol. 26, no. 4, pp. 281-289, May 2004.

- [4] S. P. Kim, *Deep Learning First Step*, Hanbit media, 2016.
- [5] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.
- [6] V. Mnih *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529-533, Feb. 2015.
- [7] I. Goodfellow, *Deep Learning*, MIT Press, 2016.
- [8] Z. Wang and L. Wang, "Adaptive negotiation agent for facilitating bi-directional energy trading between smart building and utility grid," *IEEE Transactions on Smart Grid*, vol. 4, no. 2, pp. 702-710, June 2013.
- [9] R. Y. K. Lau, M. Tang, O. Wong, S. W. Milliner, and Y. P. Chen, "An evolutionary learning approach for adaptive negotiation agents," *International Journal of Intelligent Systems*, vol. 21, no. 1, pp. 41-72, 2006.
- [10] I. Praca, C. Ramos, Z. Vale, and M. Cordeiro, "MASCeM: A multiagent system that simulates competitive electricity markets," *IEEE Transactions on Intelligent Systems*, vol. 18, no. 6, pp. 54-60, 2003.
- [11] L. Wang, Z. Wang, and R. Yang, "Intelligent multiagent control system for energy and comfort management in smart and sustainable building," *IEEE Transactions on Smart Grid*, vol. 3, no. 2, pp. 605-617, 2012.
- [12] H. Ma and H.-F. Leung, "An adaptive attitude bidding strategy for agents in continuous double auctions," *Electronic Commerce Research and Applications*, vol. 6, no. 4, pp. 383-398, Jan. 2007.
- [13] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 56, pp. 1929-1958, June 2014.
- [14] S. Ruder, "An overview of gradient descent optimization algorithm," *arXiv preprint arXiv: 1600.04747*, 2016.

저 자 소 개

이 동 구(준회원)



- 2019년 2월 : 광운대학교 전자융합 공학과 학사 졸업
- 2021년 2월 : 광운대학교 전자융합 공학과 석사 졸업
- 2021년 2월 ~ 현재 : 광운대학교 전자융합공학과 통신 및 인공지능 연구실 연구원
- 관심분야 : 인공지능, 위성통신, 스마트 그리드

이 지 영(준회원)



- 2020년 2월 : 광운대학교 전자융합 공학과 학사 졸업
- 2020년 3월 ~ 현재 : 광운대학교 전자융합공학과 석박통합과정
- 관심분야 : 인공지능, 에너지인터넷, 시계열데이터분석, 이상치 검출

경 찬 욱(준회원)



- 2017년 3월 ~ 현재 : 광운대학교 전자융합공학과
- 관심분야 : 인공지능, 디지털통신, 무선 에너지 하베스팅

김 진 영(정회원)



- 1998년 2월 : 서울대학교 전자공학과 공학박사
- 2001년 2월 : SK텔레콤 네트워크 연구소 책임연구원
- 2001년 3월 ~ 현재 : 광운대학교 전자융합공학과 교수
- 관심분야 : 차세대이동통신, 가시광통신, 전력선통신, 인공지능

※ 본 연구는 2017년도 산업통상자원부의 재원으로 한국에너지기술평가원(KETEP)의 에너지인력양성사업(No. 20194010201830)과 2021년도 광운대학교 우수연구자 지원 사업에 의해 연구되었음.