

<https://doi.org/10.7236/JIIBC.2021.21.6.163>
JIIBC 2021-6-24

심층강화학습 기반 분산형 전력 시스템에서의 수요와 공급 예측을 통한 전력 거래시스템

Power Trading System through the Prediction of Demand and Supply in Distributed Power System Based on Deep Reinforcement Learning

이승우*, 선준호*, 김수현*, 김진영**

Seongwoo Lee*, Joonho Seon*, Soo-Hyun Kim*, Jin-Young Kim**

요약 본 논문은 분산형 전력 시스템에서 심층강화학습 기반의 전력 생산 환경 및 수요와 공급을 예측하며 자원 할당 알고리즘을 적용해 전력거래 시스템 연구의 최적화된 결과를 보여준다. 전력 거래시스템에 있어서 기존의 중앙집중식 전력 시스템에서 분산형 전력 시스템으로의 패러다임 변화에 맞추어 전력거래에 있어서 공동의 이익을 추구하며 장기적인 거래의 효율을 증가시키는 전력 거래시스템의 구축을 목표로 한다. 심층강화학습의 현실적인 에너지 모델과 환경을 만들고 학습을 시키기 위해 날씨와 매달의 패턴을 분석하여 데이터를 생성하며 시뮬레이션을 진행하는 데 있어서 가우시안 잡음을 추가해 에너지 시장 모델을 구축하였다. 모의실험 결과 제안된 전력 거래시스템은 서로 협조적이며 공동의 이익을 추구하며 장기적으로 이익을 증가시킨 것을 확인하였다.

Abstract In this paper, the energy transaction system was optimized by applying a resource allocation algorithm and deep reinforcement learning in the distributed power system. The power demand and supply environment were predicted by deep reinforcement learning. We propose a system that pursues common interests in power trading and increases the efficiency of long-term power transactions in the paradigm shift from conventional centralized to distributed power systems in the power trading system. For a realistic energy simulation model and environment, we construct the energy market by learning weather and monthly patterns adding Gaussian noise. In simulation results, we confirm that the proposed power trading systems are cooperative with each other, seek common interests, and increase profits in the prolonged energy transaction.

Key Words : deep reinforcement learning, deep Q-networks (DQN), distributed power system, power transaction system, resource allocation algorithm

*준회원, 광운대학교 전자융합공학과
**정회원, 광운대학교 전자융합공학과, 교신저자
접수일자 2021년 5월 11일, 수정완료 2021년 11월 17일
계재확정일자 2021년 12월 10일

Received: 11 May, 2021 / Revised: 17 November, 2021 /
Accepted: 10 December, 2021

*Corresponding Author: jinyoung@kw.ac.kr
Dept of Electronic Convergence Engineering, Kwangwoon
University, Seoul, Korea

I. 서 론

사물인터넷(IoT: Internet of Things) 발전의 기반을 위해 에너지인터넷(IoE: Internet of Energy)은 필수적이다.

최근 글로벌 동향으로 스마트 그리드를 키워드로 각 정부에서는 신재생 에너지에 맞게 기존의 에너지시스템에서 분산형 전력 시스템을 도입하여 개인들도 전력을 거래하는 친환경적인 전력 거래시스템의 표준화 및 상용화를 앞세우며 연구가 확대되고 있다. 영국에서는 피클로(Piclo)라는 P2P 전력거래 플랫폼을 현재 상용 중이다. 피클로 플랫폼은 30분 옥션 시스템을 적용하여 신재생 에너지를 이용한 전력 공급자와 소비자를 연결해 거래하도록 하여 소비자가 DUoS (Disribution Used of System) 비용을 감소하는 효과를 가져오는 플랫폼이다. 독일에서는 소넨 커뮤니티(Sonnen Community)라는 플랫폼으로 P2P 전력거래를 진행하는데 에너지 저장 시스템을 이용해 잉여전력을 기간을 정해 거래하는 시스템이다. 그 밖에도 미국 보스턴의 옐로하(Yeloha), 미국 뉴욕 브루클린의 마이크로 그리드 샌드박스, 네덜란드의 반데브론(Vandebrom) 등에서 도입하고 있다^[1].

에너지인터넷 분야에서는 인공지능 기술을 도입하여 전력 공급 및 거래가 효율적으로 진행되는 시스템에 관한 연구가 진행되고 있다. 특히, 스마트 그리드에서는 분산형 전력 관리 시스템을 통해 에너지 자급자족이 가능한 제로 에너지빌딩을 위해 다양한 전력 시스템 알고리즘이 연구되고 있다^[2]. 심층강화학습 기반 전력 거래시스템이 활발히 연구가 진행되었으나 모델의 과적합의 문제와 실제 데이터를 통한 검증은 하는 데 있어 데이터 선택과 해석에 따른 다양한 결과를 낳는 문제가 있다. 이를 해결하기 위해 빅데이터, 심층강화학습의 최적화 알고리즘 기술들이 연구되고 있으며 모델 학습 지연시간을 감소시키는 다양한 기법들이 고안되고 있다^[3]. 기존 전력 거래시스템의 가장 큰 문제로는 서로가 이익을 과하게 추구하여 가장 이익이 되는 당사자와만 거래가 진행이 된다는 문제가 있으며 거래를 하지 못하는 대상의 경우는 지속적인 손해를 갖는 결과가 나타난다. 따라서, 본 논문에서는 분산형 전력 공급 시스템에서 공동의 이익 추구 및 협력 거래를 위해 자원 할당 알고리즘(resource allocation algorithm)을 도입하였다. 자원 할당 알고리즘은 최근 IoT에서 가전, 환경, 스마트 모빌리티 시장 등에서 확대되고 있다. 물론 많은 이익을 얻고 있는 개인의 측면에 있어서는 단기적으로는 손해를 입는다고 볼 수

있다. 그러나, 장기적으로 보게 되면 협력을 통해 자원 활용의 극대화와 거래의 효율을 높일 수 있으며 사회·경제적인 측면과 환경 보호적 측면에서의 강점의 형태로 나타난다. 본 논문에서는 자원 할당 알고리즘을 도입하여 전력 거래시스템을 구축하면 전력 생산 설비나 생산 전력을 필요한 만큼만 사용하고 잉여전력을 타인과 공유 소비를 통해 장기적인 사회의 이익 실현을 기대하고 있다. 본 논문이 제안하는 전력 거래시스템에서 전력을 거래하는 판단은 현재 상태의 전력 소비와 생산 및 날씨를 통한 미래의 수요와 공급을 예측하고 빅데이터의 학습을 통해 최적화되고 지능화된 전력 공급과 수요관리를 통해 거래가 진행된다.

본 논문은 다음과 같이 구성되어있다. II장에서는 전력 수요와 공급 데이터에 대해 분석을 하고 III장에서는 심층강화학습의 환경 설정에 관해 설명한다. IV장에서 본 논문이 제안하는 전력 거래시스템 모델에 관해 설명한다. V장에서는 DQN 시뮬레이션 결과를 확인하고 VI장에서 결론을 통해 마무리 짓는다.

II. 전력 수요와 공급 데이터 분석

1. 날씨와 월별 전력 수요와 공급에 따른 데이터 분석

분산형 전력 시스템의 대부분은 친환경 에너지(태양광 발전, 풍력 발전 등)를 사용해 전력을 공급받기 때문에 수요와 공급은 날씨의 영향을 받게 된다. 따라서, 한국 기상청의 최근 10년간의 기상 자료를 통해 마르코프 체인(Markov chain)을 통해 데이터를 분석한다. 먼저 10년간의 통계적인 날씨는 다음 그림 1과 같다. 한국 기상

Weather Statistics

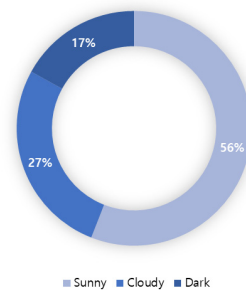


그림 1. 한국 기상 특성 분석을 통한 기상 통계.

Fig. 1. Weather statistics through Korea weather characteristic analysis.

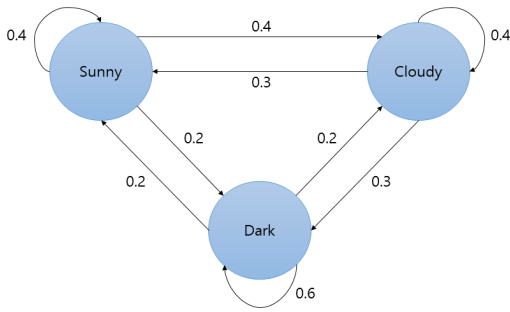


그림 2. 날씨 전이 확률.
 Fig. 2. Weather transition probability.

청의 10년간 날씨 데이터를 분석하여 현재의 기상으로부터 다음 날의 기상으로의 상태 전이 확률을 구했다. 그림 2는 이렇게 구한 상태 전이 확률을 마르코프 체인의 형태로 날씨의 변화에 대한 상태 전이 확률을 나타낸다^[4].

본 논문에서는 스마트 빌딩 단지를 하나의 분산형 전력 시스템을 이용하는 대상으로 태양광 발전 시스템이 적용되었다는 가정하에 날씨에 따른 전력 생산량을 분석하였다^[5]. 그림 3은 날씨에 따라 분석한 기상이 매우 좋을 때(맑음)와 좋지 않을 때(흐림)의 전력 발전량의 차이를 보여준다.

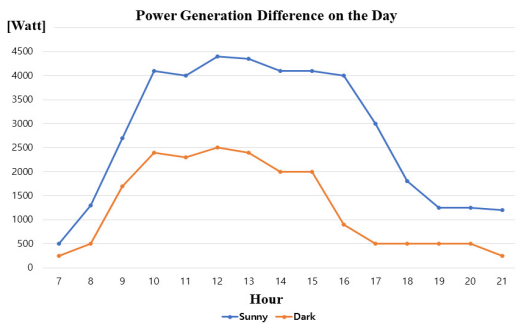


그림 3. 맑은 날씨와 흐린 날씨의 전력 발전량 차이.
 Fig. 3. Difference of power generation between the sunny and dark weather.

그림 3을 분석해 보면 흐린 날씨가 맑은 날씨에 비해 40% 정도 낮은 전력 생산량을 보이는 것을 확인할 수 있다. 즉, 신재생 에너지 공급시스템에서 전력의 수요와 공급은 날씨에 대해 높은 상관도를 보인다. 본 논문에서는 이러한 분석을 바탕으로 3가지의 날씨 상태인 맑음, 구름, 흐림에 대해 100%, 75%, 60%의 전력 생산량 효율로 데이터를 설정하였다. 표 1은 4월의 전력 수요를 기준으로 월별 전력 수요의 변화를 보여준다.

표 1. 월별 수요 변화.

Table 1. Monthly change of demand.

Month	change demand volume[%]
January	+ 6.9
February	+ 12.1
March	+ 6.9
April	0 (standard month)
May	- 1.7
June	+ 5.2
July	+ 4.8
August	+ 2.5
September	+ 0.7
October	- 4.5
November	- 4.1
December	+ 1

전력 소비량의 변화 중에서 특히, 여름철과 겨울철에 전력 수요가 많아지는 것을 확인할 수 있다^[6]. 전력 수요와 공급의 패턴을 날씨와 월별로 분석하고 가우시안 잡음을 더해 심층강화학습 환경을 구축하는 데 사용된다.

III. 심층강화학습 환경

1. 강화학습(reinforcement learning)

강화학습은 기계 학습(machine learning)의 한 영역으로 환경으로부터 정의된 에이전트(agent)가 현재의 상태를 인지하며 선택 가능한 행동 중에 보상 총합(total rewards)을 최대화하는 최적화된 정책(optimal policy)을 찾으며 학습을 하는 것이다. 즉, 정의된 에이전트가 주어진 환경과 상호작용을 하며 보상을 최대화하는 행동을 배우는 것이라고 정의할 수 있다. 다음 아래의 그림 4는 강화학습의 상호작용 과정을 보여준다.

각 단계(step)마다 에이전트와 환경은 상호작용을 하는 것을 볼 수 있다. 순서를 보면 먼저 에이전트는 환경로부터 상태(state) S_t 를 받고 에이전트는 현재의 상태에서 행동 가능한 부분에서 보상이 최대가 되는 행동(action) A_t 를 선택한다. 다음 단계($t+1$)에서 환경은 에이전트에 행동 A_t 에 대한 보상 R_t 와 상태 S_{t+1} 를 주며 보상을 최대화하는 최적화 정책(optimal policy)을 찾아가는 것이다^[7].

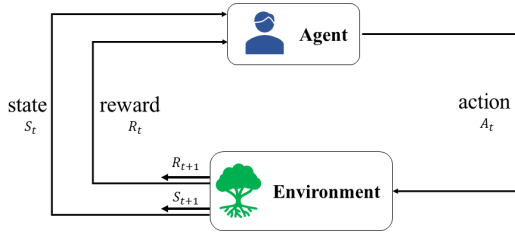


그림 4. 강화학습에서 환경 에이전트 상호작용.
Fig. 4. The environment-agent interaction in reinforcement learning.

2. Q-learning

Q-learning은 model-free 강화학습 알고리즘으로 시스템의 모델 없이 학습할 수 있는 장점이 있다. 유한한 마르코프 결정 과정(FMDP: finite Markov decision process)에서 에이전트가 특정 상태에서 특정 행동을 통해 보상을 받는 환경에서 모든 연속적인 단계들을 거쳤을 때 누적된 보상이 극대화하는 방향으로 나아가는 최적의 정책을 배우는 것이 Q-learning의 목적이다. 여기에서 학습을 위해 사용되는 Q-value는 현재 상태에서 취한 행동의 보상에 대한 가치를 상징한다^[8]. 이를 $Q(s, a)$ 라 나타내며 ‘행동-가치함수’라고 부른다. $Q(s, a)$ 를 계산할 때, 현재의 보상만이 아닌 미래까지 누적된 보상의 최상의 가치를 찾기 위해 감가율(discount factor)이 함수 계산에서 사용된다.

Q-learning은 off-policy control 방법으로 가장 큰 특징으로는 탐험을 계속하면서도 최적화된 정책을 찾을 수 있다는 장점이 있다. 또한, Q-learning은 off-policy 한 방법으로 행동 정책(action policy)과 target policy를 ϵ -greedy 알고리즘으로 결정한다. Q-learning의 $Q(s, a)$ 의 업데이트는 현재 상태 s 에서 행동 정책을 ϵ -greedy 알고리즘에 따라 ϵ 확률로 랜덤한 행동과 $1-\epsilon$ 의 확률로 최선의 행동 a 를 선택한다. $Q(s, a)$ 를 이용하여 다음 단계의 상태에서의 행동 a' 의 target policy는 다음과 같다.

$$\pi(S_{t+1}) = \operatorname{argmax}_a Q(S_{t+1}, a'). \quad (1)$$

Q-learning의 대상(target)의 ‘행동-가치함수’는 다음과 같다.

$$\hat{Q}(s, a) \leftarrow r(s, a) + \gamma \max_a \hat{Q}(s', a'). \quad (2)$$

식 (2) 을 보게 되면 대상의 ‘행동-가치함수’는 상태, 행동, 보상을 통해 도출되는 것을 확인할 수 있으며 이에

따라 Q-learning의 $Q(s, a)$ 를 업데이트하는 것은 다음과 같다.

$$Q(s, a) \leftarrow Q(s, a) + \alpha (r(s, a) + \gamma \max_{a'} \hat{Q}(s', a') - Q(s, a)). \quad (3)$$

Q-learning 알고리즘의 핵심은 이전의 값과 새로운 정보를 통한 값의 가중 합(weighted sum)을 이용하는 벨만 방정식에서 비롯된 가치 반복(value iteration) 방식을 사용한다는 것이다. Q 값은 고정된 값으로 시작하여, 에이전트의 행동 a 로 얻은 보상을 통해 갱신하고 Q-table을 이용하여 Q 값을 계산하여 Q-table에 Q 값을 업데이트한다. 이런 과정의 반복을 통해 학습이 진행되면 최적화된 정책을 찾을 수 있다. 그러나, Q-learning 알고리즘은 메모리를 많이 사용하여 복잡한 문제에 적용하기 어렵다는 단점이 있다. 또한, 특정 조건에서 행동 가치를 과대평가한다는 문제를 가지고 있다.

2. DQN(Deep Q-Networks)

Q-learning 알고리즘의 메모리 부족 문제를 해결하기 위해 신경망 네트워크(NN: Neural Network)를 통한 deep Q-learning이 고안되었다. DQN은 상태와 행동에 대한 보상의 값을 Q-table에 저장하지 않고, 신경망 네트워크를 구축하여 상태와 행동에 대한 보상을 얻는 학습이다. 이때, experience replay를 이용하고 심층 신경망 네트워크를 사용한 것을 DQN 알고리즘이다.

Algorithm Deep Q-Networks (DQN)

```

Initialize primary network  $Q_\theta$ , target network  $Q_{\theta'}$ 
Initialize replay memory  $\mathcal{M}$  with random experiences to capacity  $C$ 
for episode, E do
    initialize Environment parameters
    while episode is not done do
        for each environment step, S do
            Observe state  $s_t$ 
            With probability  $< \epsilon$  select random action  $a_t$ 
            else select  $a_t = \operatorname{argmax}_a Q^*(\phi(s_t), a; \theta)$ 
            Execute  $a_t$  and observe next state  $s_{t+1}$  and reward  $r_t$ 
            Set  $S_{t+1} = s_t, a_t, r_t$  and  $\phi_{t+1} = \phi(S_{t+1})$ 
            Store  $(s_{t+1}, a_t, r_t, s_t)$  in  $\mathcal{M}$ 
            Randomly sample  $j$  from  $\mathcal{M}$  with minibatch size  $S$  of
            Experiences  $(s_{j+1}, a_j, r_j, s_j)$ 
            Set  $y_j = \begin{cases} r_j & \text{(for terminal)} \\ r_j + \gamma \max_a Q(\phi_{j+1}, a; \theta) & \text{(for non-terminal)} \end{cases}$ 
            Perform a gradient descent step
        end For
        perform learning update target  $Q$  network for weights  $\theta$ 
    end while
end For
    
```

그림 5. Deep Q-Networks 알고리즘 체계.
Fig. 5. Deep Q-Networks algorithm scheme.

또한, DQN은 상태 간의 상관도가 크다는 문제를 해결하고자 experience replay를 이용해 새로운 경험을 바로 학습하는 것이 아니라 experience memory에 저장한 후에 replay를 이용해 학습을 진행한다^[9]. DQN은 현재 상태에서의 Q 값과 비교해야 할 target 값에서 사용되는 다음 상태에서의 Q 값은 같은 θ (theta)를 사용한다. 그림 5는 experience replay를 통해 최적의 정책을 찾는 DQN의 알고리즘을 나타낸다^[10].

IV. 전력 거래시스템 모델

본 논문은 심층강화학습 기반 분산형 전력거래를 위해 자원 할당 분배 알고리즘을 통한 전력 거래방식을 제안한다. 그림 6은 본 논문에서 제안하는 전력 거래시스템의 DQN 알고리즘의 구조를 나타낸다. 에피소드를 계속 반복을 하며 최적의 target Q를 통해 최적의 전력 거래 시스템을 학습한다. 이렇게 학습이 된 모델을 통해 모의 실험을 함으로써 성능을 비교하기 위해 전력거래 총량, 중앙 전력공급시스템의 의존도, 전력 가격의 적정성 및 전력 보유량을 시뮬레이션 결과에서 분석하고자 한다^[11-12].

분산형 전력 공급 시스템의 전력거래 방법은 다음 식 (4)~(10)을 통해 진행된다. 이때, 에이전트는 스마트 빌딩을 대상으로 얻은 전력 데이터 기반으로 한다^[13]. 각 에이전트의 현재의 잉여전력은 다음과 같다.

$$\begin{aligned} \text{remain power}_t &= \text{powersupply}_t \\ &\quad - \text{power demand}_t \\ &\quad + \text{remain power}_{t-1}. \end{aligned} \quad (4)$$

식 (4)을 통해 각 에이전트의 데이터로는 현재의 전력 수요와 공급량을 통해 현재의 잉여전력을 판단한다. 미래의 예측되는 잉여전력은 다음의 식을 통해 구해진다.

$$\begin{aligned} \text{remain power}_{\text{predict}} \\ = \text{remain power}_t \times \text{pattern}_{\text{weather, mon}} \end{aligned} \quad (5)$$

식 (5)에서는 거래하는 달의 정보와 내일의 일기예보 정보의 패턴을 분석하여 내일의 잉여전력을 예측한다. 이렇게 구한 예측되는 잉여전력을 통해 현재 거래 가능하다고 예측되는 전력량은 다음과 같다.

$$\begin{aligned} \text{tradeable power}_{\text{predict}} \\ = \text{remain power}_{\text{predict}} \\ \times \text{weather}_{\text{probability}} \\ \times \text{monthly}_{\text{demand pattern}}. \end{aligned} \quad (6)$$

식 (6)에서의 전력거래 가능한 양의 결정은 예측되는 전력량, 날씨에 따른 전력 생산, 매월의 수요 패턴 인자들의 곱으로 이루어진다. 현재의 잉여전력에서 날씨를 통한 전력 공급량과 월별 수요의 변화에 대한 분석을 바탕으로 예측되는 잉여전력을 통해 전력 거래가능량을 제시한다^{[6][14]}. 이렇게 구한 판매자와 소비자의 전력거래 가능한 양을 통해 자원 할당 알고리즘을 적용하기 위해 다음의 식을 이용한다.

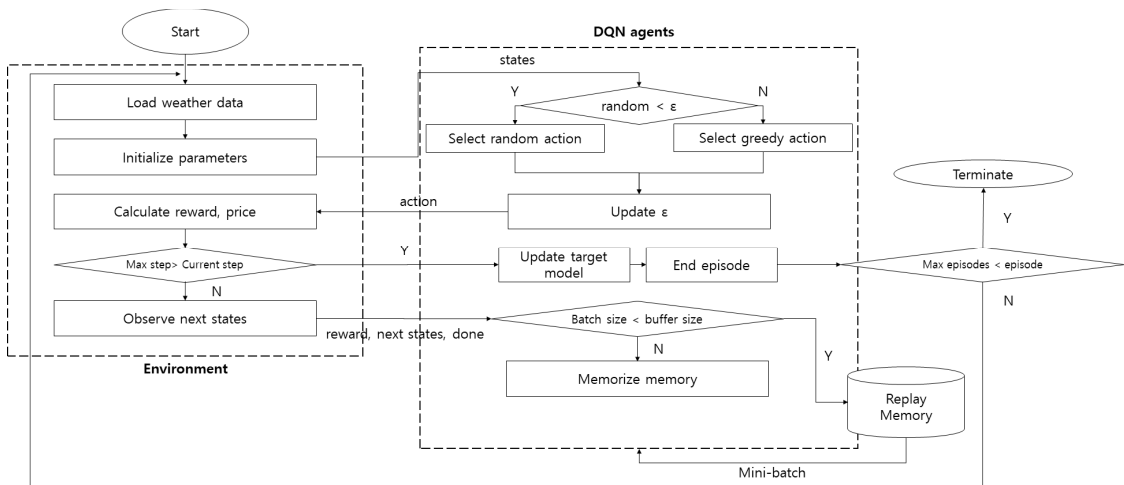


그림 6. 제안된 전력 거래시스템의 DQN 알고리즘 구조.

Fig. 6. DQN algorithm structure in the proposed power transaction system.

$$\begin{aligned} & \text{average tradeable power} \\ & = \text{avg}(\text{tradeablepower}_{\text{predict}}^{\text{seller}}, \text{tradeablepower}_{\text{predict}}^{\text{consumer}}) \end{aligned} \quad (7)$$

식 (7) 을 통해 판매자와 소비자의 거래가 가능한 전력량의 평균을 구한다.

$$\begin{aligned} & \text{transaction power}_{\text{real}} \\ & = \max(\text{tradeablepower}_{\text{predict}}^{\text{seller}}, \text{tradeablepower}_{\text{predict}}^{\text{consumer}}) \\ & \quad - \text{average tradeable power}. \end{aligned} \quad (8)$$

실제 전력 거래가 되는 양은 식 (8) 을 통해 진행된다. 각 에이전트의 전력거래 가능한 양으로 예측되는 식 (6) 을 통해 구한 값에서의 최댓값을 구하고, 식 (7) 에서 구한 평균 전력 거래가능량 값을 뺀 전력량을 통해 거래가 진행이 된다.

$$\begin{aligned} & \text{price variation} \\ & = \text{principal}_{\text{price}} \times (1 + \text{price}_{\text{ratio}}). \end{aligned} \quad (9)$$

식 (9) 을 통해 전력거래를 하기 전에 각 에이전트의 이익의 영향을 받도록 전력거래 가격의 원금에서 전력거래를 통한 예상되는 이익과 전력량 등의 정보를 통한 가격 비율을 통해 전력 가격이 변동되도록 하였다.

$$\begin{aligned} & \text{transaction price}_t \\ & = \text{price variation} \\ & \quad \times (1 + \beta \times \text{ratio}_{\text{remainpower}}). \end{aligned} \quad (10)$$

식 (10) 에서는 최종 전력거래 가격의 책정을 한다. 최종 거래 가격은 발전량과 공급량의 비율과 날씨, 매월의 패턴의 비율의 곱으로 결정하였다.

V. DQN 시뮬레이션 결과

그림 7은 주 단위 배터리의 평균적인 전력 보유량의 시뮬레이션 결과를 보여준다. 결과를 보게 되면 거래 참가자들이 전력 자원 할당 알고리즘을 통한 전력거래를 해서 결과가 비슷한 양상을 보인다. 또한 서로 전력이 부족할 때도 전력을 거래하여 중앙집중식 전력공급원으로부터 즉각적으로 전력을 공급을 받지 않고 협력하는 양상을 보여준다. 그림 8은 주 단위 평균적인 전력거래 가격을 보여준다. 서로 협력하며 전력을 거래하지만 서로 이익을 추구하기 위해 전력거래의 가격을 결정하는 데는 자신에게 이익이 되면서 공동의 이익을 위해 적절한 가격 산정하는 것을 볼 수 있다.

그림 9는 하루 단위 배터리의 전력 보유량의 시뮬레이션

결과를 보여준다. 결과를 보게 되면 다소 변동이 심한 것을 볼 수 있다. 이러한 변화는 매일 전력을 거래하고 있다는 것을 나타낸다. 그림 10은 하루 단위의 전력거래 가격을 보여준다. 현재의 에이전트마다 정보에 따라 가격의 변동이 있는 것을 볼 수 있다. 시뮬레이션 결과 일정 가격 이상은 올라가지 않고 전력거래를 하는 데 있어서 협력하며 전력을 거래하는 안정된 전력 거래시스템임을 확인할 수 있다.

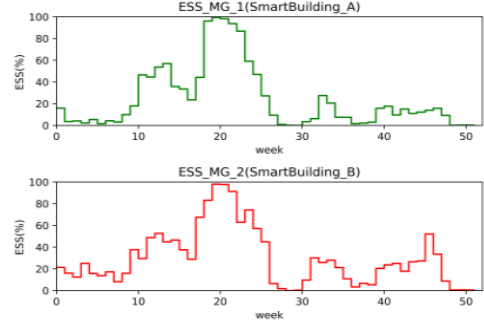


그림 7. 주간 평균 배터리 수준.
Fig. 7. Weekly average battery level.

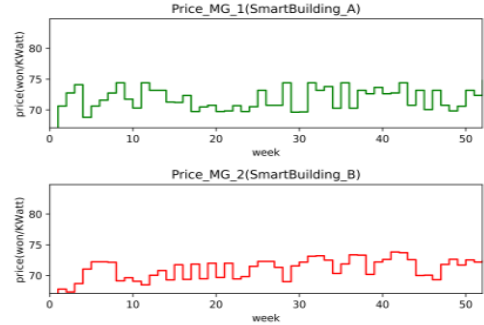


그림 8. 주간 평균 전력 거래 가격.
Fig. 8. Weekly average power trading price.

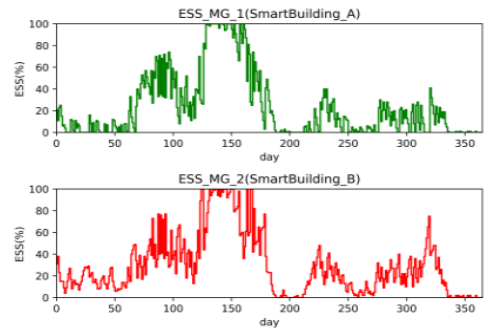


그림 9. 일일 평균 배터리 수준.
Fig. 9. Daily average battery level.



그림 10. 일일 평균 전력 거래 가격.
 Fig. 10. Daily average power trading price.

그림 11은 시뮬레이션 결과에서 에피소드 단위의 전력 거래량을 보여준다. 결과를 보게 되면 에피소드가 증가할수록 전력 거래량은 증가하며 이익을 최대화하는 것을 볼 수 있다. 전력거래를 하는 데 있어서 에이전트는 협력하며 전력 자원을 할당해서 거래하는 것이 장기적인 이익이라는 것을 학습했다고 보인다. 그림 12는 이렇게 학습된 에피소드별 점수로 환산을 했을 때 초반보다 많이 증가하여 140대의 근처에서 머무르는 것을 볼 수 있다.

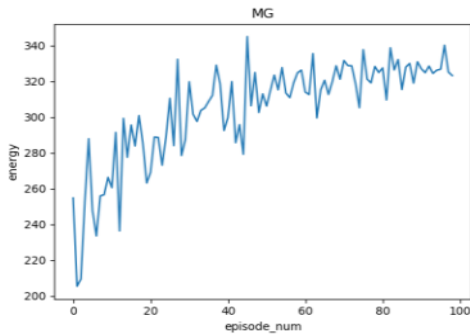


그림 11. 에피소드당 전력 거래량.
 Fig. 11. Power transaction volume per episode.

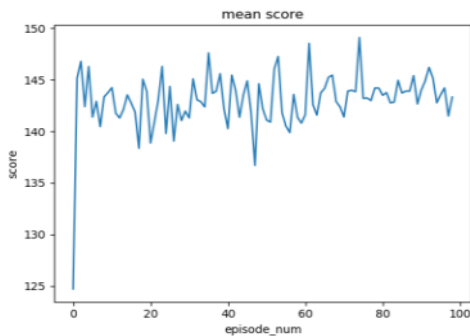


그림 12. DQN에서 에피소드당 평균 점수.
 Fig. 12. Mean score per episode in DQN.

그림 13, 14는 시뮬레이션 결과에서 에피소드 단위별 중앙 전력 공급 시스템의 의존도를 보여주지 위한 지표들이다. 실제로 분산형 전력 거래시스템이 운용되더라도 중앙 전력 공급 시스템은 필수적이다. 그러한 이유는 신재생 에너지의 전력 공급의 불안정성 때문이다. 그러나, 이러한 중앙 전력 공급 시스템의 부담이 높아지면 사회 전체 전력 공급 시스템의 문제를 초래할 수 있다. 따라서, 존재하는 중앙 전력 공급 시스템에 덜 의존적인 분산형 전력 공급 시스템을 위해 서로 협력하며 전력거래를 해야 한다. 시뮬레이션 결과 중앙집중식 전력 공급 시스템의 의존도가 20% 이하로 낮은 것을 확인할 수 있다. 이는 중앙집중식 전력 공급 시스템의 부담을 줄여주는 전력 거래시스템임을 확인할 수 있다.

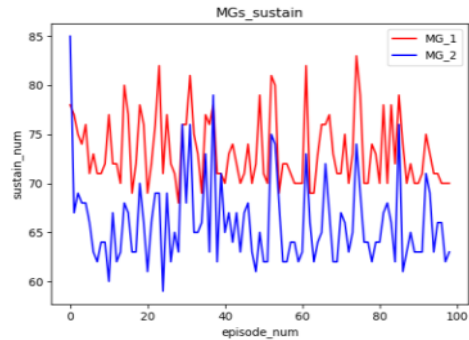


그림 13. 에피소드당 전력 유지 횟수.
 Fig. 13. The number of power sustain per episode.

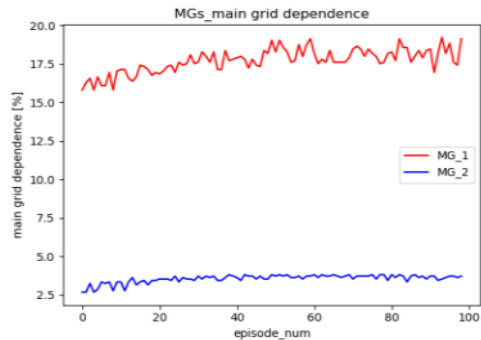


그림 14. 메인 그리드에 대한 에이전트의 전력 의존율.
 Fig. 14. The percentage of the agent power dependence on the main grid.

VI. 결 론

본 논문에서는 분산형 전력 거래시스템 모델과 DQN

알고리즘에 관해 설명하고 전력거래 시스템에서 미래의 날씨와 매월의 패턴을 예측하여 전력거래를 하는 데 있어서 효율적인 전력거래를 하며 자원을 할당하여 중앙집중식 전력 공급시스템의 부담을 완화하기 위해 DQN 알고리즘을 적용한 전력 거래시스템을 제안하였다. 제안한 시스템을 실제 전력시장의 데이터를 바탕으로 유사한 환경을 만들어 시뮬레이션을 진행하였고 결과를 확인해보면 에이전트가 협력하며 전력을 거래하고 기존의 중앙집중식 전력 시스템의 의존도 감소 면에서 우수한 성능을 보이는 것을 확인하였다. 또한, 자원 할당 알고리즘을 통해 에이전트가 자신의 행동에 따른 미래의 상태에 대한 예측을 통해 이익이 최대가 되도록 전력 거래량을 증가시켜 거래하는 것을 증명하였다. 그러나 현실에서의 분산형 전력 공급 시스템의 전력 생산과 소비에 경우 변수가 다양하고 구현 관점에서는 현재로서는 한계점이 있다. 제안한 시스템 모델을 실제로 구현하기 위해서는 심층 강화학습에서의 많은 연산량을 시간지연 없이 처리할 수 있는 초저지연 처리장치 및 DQN 모델의 많은 파라미터를 저장하고 있을 저장장치가 필요하다. 또한 데이터를 학습하는 데 있어서 상관도 있는 데이터를 분류하고 신뢰성 있는 빅데이터를 학습하여야 한다는 한계가 있다. 하지만 현재의 빅데이터와 심층강화학습 기술로 시간지연 없이 처리할 수 있는 획기적인 알고리즘 구현하는 것은 한계가 있어 현실적으로 현재의 전력 시스템 적용은 무리이며 향후 심층강화학습의 기술이 발전된다면 구현이 가능해질 것이다. 그러므로 현실의 전력 거래시스템에 적용을 위해 상관도 있는 빅 데이터와 초저지연이 가능한 획기적인 알고리즘 연구와 다수 사용자에게 맞는 전력 거래시스템인 분산형 전력거래 맞춤형 시스템의 연구가 진행될 계획이다.

References

- [1] C. K. Park, "A study on P2P power trading possibilities in Korea", *The Magazine of the Korea Energy Economics Institute*, pp. 3-29, April, 2016.
- [2] X. Lu, X. Xiao, L. Xiao, C. Dai, M. Peng, and H. V. Poor, "Reinforcement learning-based microgrid energy trading with a reduced power plant schedule," *IEEE Internet of Things Journal*, Vol. 6, No. 6, pp. 10728-10737, Dec. 2019, doi: 10.1109/JIOT.2019.2941498.
- [3] T. Levent, P. Preux, E. le Pennec, J. Badosa, G. Henri and Y. Bonnassieux, "Energy Management for Microgrids: a Reinforcement Learning Approach," *2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, pp. 1-5, Nov. 2019. Doi: 10.1109/ISGTEurope.2019.8905538.
- [4] KMA, "Number of days by rank of all clouds," Accessed: 2020. [Online]. Available: <https://data.kma.go.kr/stcs/grnd/selectCloudSpecifyDay.do?pgmNo=671>
- [5] AccuWeather. Accessed: 2020. [Online]. Available: <https://www.accuweather.com/ko/kr/seoul/226081/july-weather/226081>
- [6] S. J. Cho and T. Y. Yoon, "Analysis and implications of seasonal pattern for residential electricity demand," *Korea Energy Economics Institute (KEEI), Energy Focus*, Vol. 13, No. 2, 2016.
- [7] M. van Otterlo and M. Wiering, "Reinforcement learning and Markov decision processes," *Springer, Berlin, Heidelberg*, Vol. 12, 2012. Doi: org/10.1007/978-3-642-27645-3_1
- [8] Chris Watkins, "Learning from delayed rewards", *Doctoral Dissertation*, King's College. Jan. 1989.
- [9] S. Y. Jang *et al.*, "Research trends on deep reinforcement learning," *The magazine of the Electronics and Telecommunications Trends*, ETRI, Vol. 34, No. 4, 2019. DOI: 10.22648/ETRI.2019.J.340401
- [10] Mnih V, Kavukcuoglu K, Silver D, *et al.*, "Playing Atari with deep reinforcement learning," *Google Deepmind Technologies*, Dec. 2013. Available from: <http://arxiv.org/abs/1312.5602>
- [11] H. S. Oh and G. S. Yu, "Mechanical-electrochemical-heat transfer multiple physics analysis techniques for secondary battery thermal behavior analysis," *Gwangju Institute of Science and Technology (GIST), Latest Research Trends of ESS*, Vol. 60, No. 1, 2020.
- [12] M. Kim, S. Parkt, J. K. Choi, and J. Lee, "Energy independence of energy trading system in microgrid," *2017 IEEE Innovative Smart Grid Technologies - Asia (ISGT-Asia)*, pp. 1-4, 2017. Doi: 10.1109/ISGT-Asia.2017.8378441.
- [13] S. I. Kim and S. G. Kim, "Smart building that gets smarter: trends and examples of smart buildings," *KT Economic Management Research Institute, Digieco report*, 2017.
- [14] KEPCO, "Power consumption by industry classification". [Online]. Available: <https://bigdata.kepcoco.kr/cmsmain.do?scode=S01&pcode=000167&redirect=Y>. Dec. 2020.
- [15] P. J. Kwon, H. D. Lee, D. H. Tae, J. H. Park, M. Ferreira, and D. S. Rho, "Operation method of power supply for Eco-friendly movable-weir based on natural energy sources," *Journal of the Korea Academia-Industrial cooperation Society*, Vol. 21, No. 2, pp. 601-610, 2020.

doi: <https://doi.org/10.5762/KAIS.2020.21.2.601>

- [16] S. J. Lee, B. W. On, G. S. Choi, and S. Yi, "DQN model based on virtual environment for improving high score convergence speed in reinforcement learning games," *The Proceedings of the 2020 Korean Institute of Information Technology (KIIT) Conference*, pp. 472-475, 2020.

김진영(정회원)



- 1998년 2월 : 서울대학교 전자공학과 공학박사
- 2001년 2월 : SK텔레콤 네트워크 연구소 책임연구원
- 2001년 3월 ~ 현재 : 광운대학교 전자융합공학과 교수

- 관심 분야 : 차세대이동통신, 가시광통신, 전력선통신, 인공지능

저자 소개

이승우(준회원)



- 2021년 2월 : 광운대학교 전자융합공학 학사 졸업
- 2021년 3월~현재 : 광운대학교 전자융합공학과 석박사통합과정
- 관심 분야 : 인공지능, 스마트 그리드, 무선 에너지하베스팅, 디지털통신

선준호(준회원)



- 2021년 2월 : 광운대학교 전자융합공학 학사 졸업
- 2021년 3월~현재 : 광운대학교 전자융합공학과 석박사통합과정
- 관심 분야 : 인공지능, 무선 에너지하베스팅, 디지털통신

김수현(준회원)



- 2019년 2월 : 광운대학교 전자융합공학과 졸업
- 2019년 3월~ 현재 : 광운대학교 전자융합공학과 석박사통합과정
- 관심 분야 : 인공지능, 전력선통신, 스마트 그리드

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