ISSN 1975-8359 [Print] / ISSN 2287-4364 [Online]
The Transactions of the Korean Institute of Electrical Engineers, vol. 69, no. 11, pp. 1659~1667, 2020
https://doi.org/10.5370/KIEE.2020.69.11.1659

Distribution Network Reconfiguration to Minimize Power Loss Using Deep Reinforcement Learning

전력손실 최소화를 위한 심층 강화학습 기반 배전계통 재구성

Se-Heon Lim · Tae-Geun Kim · Sung-Guk Yoon

임세헌* · 김태근* · 윤성국†

Abstract

Distribution network reconfiguration (DNR) is a technique that changes the status of sectionalizing and tie switches for various purposes such as loss minimization, voltage profile improvement, load leveling, and hosting capacity increase. Although previous algorithms for DNR show good performance, they still have practical limitations. Most of the algorithms assumed that a central coordinator knows all parameters and/or perfect states in a distribution network. Reinforcement learning which is a model-free optimization technique can be a key way to overcome these limitations. This work proposes a DNR scheme using deep reinforcement learning to minimize power loss defined by the amount of line loss and renewable energy curtailment. We model the DNR problem as a Markov decision process (MDP) problem and apply the reinforcement learning algorithm to solve this problem in real-time. Simulation result using 33-bus radial distribution system shows that the proposed scheme shows similar performance compared to an existing method which uses all information on the distribution network.

Key Words

deep Q network (DQN), distribution netwok reconfiguration (DNR), line loss, reinforcement learning, renewable energy curtailment

1. 서 론

지속가능한 성장을 위해 전 세계 각국은 재생에너지의 비중을 늘리고 있다. 우리나라도 태양광과 풍력을 중심으로 하여 2030년까지 재생에너지 설비용량 63.8 GW 보급을 목표로 재생에너지 비율을 20%까지 높이는 '에너지 3020정책'을 시행중에 있다[1]. 같은 취지로 정부는 1 MW 이하 소규모 재생에너지 발전원의 배전망 접속을 보장하는 정책을 시행하였다[2]. 그러나 불확실성 전원인 재생에너지가 다량으로 계통에 연계되면 역조류 현상, 과전압, 하모닉스 등을 일으켜 전력품질에 악영향을 미치게 된다[3]. 따라서 재생에너지는 계통의 안정성을 유지하는 한도 내에서 계통으로의 유입이 허가돼야 하고 더 많은 재생에너지를 수용하기 위해서는 변전설비와 배전망 등 공용 전력망을 보강해야 한다.

전력망이 다량의 재생에너지를 수용하기 위해서는 장기적인 관점으로 배전설비의 보강은 피할 수 없으나 이는 시간 및 비용이 많이 소모되기 때문에 현재 있는 전력설비인 OLTC(On-Load Tap Changer), 인버터, ESS, 개폐기 등을 최대한 활용하는 방안을 고려해야한다[4-6]. 이러한 전력설비의 활용은 전력망의 보강 없이 다수의 재생에너지가 존재하는 상황에서 계통의 안정성을 유지하기 위해 감수하는 재생에너지의 출력삭감[7]을 최소화하여 전 지구적인 에너지 손실을 방지할 수 있다.

배전계통의 전력설비 중 절체용 개폐기는 방사형 계통에서 피더의 말단을 연결하는 개폐기로 평소에는 개방되어 있다가 계통에 사고가 발생하면 전력이 차단된 말단 모선을 근처의 정상 계통으로 절체하여 사고에 대처하는 개폐기이다. 개폐기 동작에 의해 계통의 토폴로지가 변경되면 그에 따라 전력조류가 바뀌는 효과가 있다. 기존 연구에서는 이러한 특성을 활용하여 배전계통 재구성을 통해 고장뿐만 아니라 선로손실 감소[8], 수용용량 증대[7] 및 전압안정도 향상[9] 등 다양한 목적으로 운영이 가능함을 확인하였다. 기존 연구에서 행해졌던 배전계통 재구성 기법은 크게 두 부류로 구분된다. 먼저 배전망 및 모든 상태 정보를 안다는 가정 하에 이상적인 해를 구하는 최적화 기법과 과거 데이터로 축적된 경험을 기반으로 하는 휴리스틱 기법이 있다. 최적화 기법은 혼합 정수 선형계획법[10], 동적 계획법[11], 확률적 혼합정수 선형 계획법[12] 등으로 분류될 수 있다. 두 번째 기법인 휴리스틱 기법은 최소 스패닝 트리로 배전계통 재구성 문제를 푸는 데 이용되었

† Corresponding Author : Dept. of Electrical Engineering, Soongsil University, Korea.
E-mail: sgyoon@ssu.ac.kr
https://orcid.org/0000-0002-8987-6628
* Dept. of Electrical Engineering, Soongsil University, Korea.
https://orcid.org/0000-0001-7049-4163 https://orcid.org/0000-0002-6676-4869
Received : July 4, 2020 **Revised :** October 8, 2020 **Accepted :** October 24, 2020

다[13].

기존의 기법들 중 최적화 기법은 배전계통 재구성에 적용되어 성능 향상을 확인했음에도 계속해서 변동하는 미래의 전력정보를 정확하게 안다는 것은 불가능하기 때문에 실제적으로 정확한 제어가 어렵다는 한계를 가지고 있다. 휴리스틱 기법의 경우에는 문제를 총체적으로 해결하는 것이 아니라 근시안적으로 판단을 내리기 때문에 최적해와 차이가 날 수 있다. 본 논문에서는 배전계통 재구성을 Markov decision process(MDP) 문제로 모델링하고 배전계통 재구성의 상태전이 함수를 알지 못하더라도 학습을 통하여 환경을 스스로 이해하고 상황에 맞는 판단을 내리는 강화학습 기반 배전계통 재구성 기법을 제안한다.

강화학습은 의사결정 방식으로 분류되는 게임, 로봇, 자율주행 등의 분야에서 인간을 뛰어넘는 성과를 내기도 했다[14-16]. 전력 분야에서도 강화학습의 강점으로 의사결정 문제인 MDP로 표현 가능한 에너지 스케줄링[17], 수요반응[18], 전압 안정도[19] 등의 분야에서 효용성이 검증되었으며 본 논문에서 제안하는 배전계통 재구성에서도 활용된 바 있다[20].

변동성과 불확실성이 높은 실제 환경에서 한계를 보였던 기존 기법들에 비해 제안하는 강화학습 기반 기법은 데이터를 통해 현 상태에 맞는 해결책을 제시하는 방식으로 실질적인 적용 가능성이 높다고 평가된다. 본 논문에서는 전력손실을 선로손실과 재생에너지 출력삭감의 합으로 정의하고 이를 최소화하는 방향으로 강화학습을 학습시킨다. 제안하는 강화학습 기반 배전계통 재구성 기법은 배전계통 33 테스트 모선의 사례연구를 통해 전역 정보를 가지고 최적화를 수행하는 기법과 거의 유사한 성능을 보이는 것을 확인하였다.

2. 심층 강화학습 (Deep Reinforcement Learning)

강화학습은 기계학습 기법 중 하나로, 행동의 주체인 에이전트(agent)가 주어진 상태(state)에서 어떤 행동(action)을 취할 것인지를 보상(reward)을 통해 스스로 학습하는 알고리즘이다. 에이전트는 행동을 할 때마다 보상을 받게 되고 누적 보상을 최대화하는 방향으로 학습을 진행한다. 따라서 강화학습 설계자는 문제의 목적에 적합한 보상 함수를 설계함으로 에이전트가 문제를 효율적이고 정확하게 학습하도록 유도한다. 상태에 따라 어떤 행동을 할 것인지 정의한 것을 정책(policy)이라고 하는데 강화학습은 정책을 학습하는 방법에 따라 오프폴리시(off-policy)와 온폴리시(on-policy) 알고리즘으로 분류된다. 온폴리시는 학습하는 정책과 행동하는 정책이 같은 경우이며, 오프폴리시는 학습하는 정책과 행동하는 정책이 다른 경우로 대표적으로 Q-러닝이 있다. 온폴리시에 비해 오프폴리시가 갖는 장점은 과거 데이터를 가지고 학습이 가능하다는 점과 온폴리시에 비해 샘플링 효율이 높다는 점 등이 있다[21]. 본 논문에서는 오프폴리시의 대표적인 기법인 Q-러닝을 사용하였다. 모델정보 없이 학습이 가능한 Q-러닝은 행동 가치 함수인 $Q(s^t,a^t)$를 학습하여 최적의 정책을 찾기 위해 사용된다. Q-러닝의 학습은 식(1)과 같은 반복 과정을 통해 이루어지는데 Q 함수가 수렴하면 식(2)에서처럼 수렴된 Q 함수를 기반으로 주어진 상태에서 가능한 행동 중에 $Q(s^t,a^t)$값이 최대인 행동을 결정하여 최적의 정책(π^*)으로 수행하게 된다.

$$Q(s^t,\ a^t) \leftarrow Q(s^t,\ a^t) + \alpha[r^{t+1} + \gamma \cdot \max_a Q(s^{t+1},\ a) - Q(s^t,\ a^t)] \quad (1)$$

$$\pi^* : s^t \rightarrow \arg\max_a Q^*(s^t,\ a) \quad (2)$$

여기서 s^t, a^t, r^t는 시간 t에서의 상태, 행동, 보상을 의미하며 α, γ는 각각 학습율과 감가율을 나타낸다. 그러나 위와 같은 단순한 Q-러닝은 Q 함수를 상태와 행동을 축으로 하는 행렬로 구성하여 업데이트하는 방식을 취하기 때문에, 상태나 행동의 경우의 수가 아주 많아지면, 기하급수로 늘어나는 계산량에 의해 현실적으로 사용이 불가능하다. 이를 해결하기 위해 Q-러닝에 지도학습인 딥러닝(Deep Learning) 기법을 접목한 Deep Q-Network(DQN)이 개발되어 심층 강화학습의 주요기법으로 활용되고 있다. DQN은 Q값을 구하기 위하여 행렬 형태로 값을 저장하여 업데이트 하는 방식 대신에 입력을 상태로 하고 출력을 Q값으로 하는 인공신경망의 파라미터를 업데이트 하여 Q 함수를 근사하는 방식이다. DQN의 파라미터(θ^Q)는 식(3)의 Q-목표와 Q값의 오차제곱을 최소화하는 방향으로 학습함으로 연속적인 Q 함수를 근사한다. Q 함수를 학습시키는 데이터로 식(4)와 같이 상태, 행동, 보상, 다음 상태의 경험튜플을 이용한다[22].

$$L(\theta^Q) = [r^t + \gamma \bullet \max_{a'} Q(s^{t+1},\ a';\theta^{Q-}) - Q(s^t,\ a^t;\theta^Q)]^2 \quad (3)$$

$$e^t = (s^t,\ a^t,\ r^t,\ s^{t+1}) \quad (4)$$

3. 전력손실 최소화를 위한 목적함수 설정

본 논문에서는 전력손실을 선로손실과 재생에너지의 출력삭감으로 정의하고 이를 최소화하는 기법을 제안한다. 재생에너지의 출력삭감도 전력손실로 간주한 이유는 다수 재생에너지원이 도입된 상황에서 출력삭감은 상당한 손실로 경제성에 영향을 끼치기 때문이다. 본 논문에서는 재생에너지가 계통의 안정성을 나타내는 기준인 제약조건(식(5)-식(6))을 만족하는 한도 내에서 계통으로의 연계가 가능하다고 가정하였다.

$$V_{\min} < v_n^t < V_{\max} \quad \forall n \in N \quad (5)$$

$$i_e^t < I_{\max} \quad \forall e \in E \quad (6)$$

여기서 v_n^t는 시간 t에서 모선 n의 전압, $V_{\min}$ 및 $V_{\max}$은 각각 전압 유지범위의 최소, 최대전압, i_e^t는 시간 t에서 선로 e에 흐르는 전류이며 $I_{\max}$는 허용전류를 의미하고 N과 E는 모선집합과 선로집합을 나타낸다. 본 논문에서는 $\overline{n}$을 태양광 발전기가 설치된 모선으로 정의하고 식 (5), (6)의 제약조건을 만족하면서 모선 $\overline{n}$에서 배전망으로 연계 가능한 최대 전력을 제한전력($p_{\overline{n}}^t$)이라고 정의하였다. 출력삭감($c_{\overline{n}}^t$)은 전압 유지범위를 만족시키기 위해 재생에너지의 출력전력에서 삭감되는 전력으로, 식(7)과 같이 발전기 출력에서 계통연계 제한전력을 제하여 구한다. 배전계통에서 부하전력과 발전전력이 실시간으로 변동하므로 모선 $\overline{n}$에 설치된 발전기의 계통연계 제한전력($p_{\overline{n}}^t$) 또한 시간 t 에 대한 함수로 표현된다.

$$c_{\overline{n}}^t = \max(P_{\overline{n}} \cdot \rho^t - p_{\overline{n}}^t,\ 0) \tag{7}$$

여기서 $P_{\overline{n}}$은 모선 $\overline{n}$에 설치된 발전기 용량이고, ρ_t는 시간 t 에서의 태양광 발전효율을 의미한다. 본 논문에서는 재생에너지원으로 일반 배전망에 주로 설치되는 태양광 발전원만을 고려한다. 시간 t에 따른 태양광 발전효율 ρ_t은 계절 및 시간에 따른 태양의 고도와 비례하는 성질을 이용하여 태양의 고도가 90도일 때를 최대 출력으로 가정하여, 식(8)에서와 같이 정의하였다[23].

$$\rho_t = \max(\cos(\theta) \cdot \cos(\Phi) \cdot \cos(15^\circ(t-12)) + \sin(\theta) \cdot \sin(\Phi),\ 0) \tag{8}$$

여기서 θ는 위도, Φ는 적위를 의미한다. 본 논문에서는 전력손실을 제어주기(T) 동안의 출력삭감($c_{\overline{n}}^t$), 선로손실(l^t) 및 개폐기 동작에 의한 개폐 손실(sl^t)의 합으로 정의하고 이에 따라 목적함수를 식 (9)과 같이 정의하였다.

$$f = \sum_{t=1}^{T} (c_{\overline{n}}^t + l^t) \cdot \triangle t + sl^t \tag{9}$$

본 연구에서는 개폐기의 개/폐 구성에 따라 계통연계 제한전력($p_{\overline{n}}^t$)과 선로손실(l^t)이 달라지는 특성을 이용하여 현 계통 상황에서 전력손실 최소화를 위해서 목적함수 f를 최소화하는 절체용 개폐기의 개/폐 구성을 결정한다. 다음 장에서는 배전계통 재구성을 MDP 문제로 모델링하여 강화학습을 통한 해법을 제시한다.

4. 배전계통 재구성 문제 Markov decision process(MDP) 모델링

강화학습은 일반적으로 의사결정과정이 확률과 그래프로 표현되는 MDP 문제의 해법으로 활용된다. MDP는 그림 1 과 같이 대표적으로 상태(state), 행동(action), 보상(reward)으로 구성된다. 배전계통 재구성은 MDP 문제로 정의될 수 있으며, 4.1-4.3에서는 배전계통 재구성 문제를 MDP로 모델링하여 각 구성요소에 관해 설명하고 4.4에서는 핵심기술인 심층 강화학습기반 배전계통 재구성 알고리즘을 기술한다. 4.5에서는 4.4의 알고리즘으로 학습된 DQN의 한계점을 개선하기 위해 보정하는 알고리즘을 추가로 제안한다. 제안하는 연구의 상태, 행동은 배전계통 재구성 선행연구 [20]의 것을 사용하며, 보상과 보정 알고리즘은 본 연구의 독창적인 부분이다.

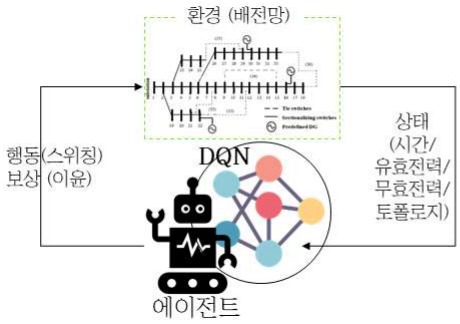


그림 1 강화학습 기반 배전계통 재구성 개념도
Fig. 1 The concept of reinforcement learning based distribution network reconfiguration

4.1 행동(action)

MDP에서 행동 집합이란 주어진 상태에서 학습 주체인 에이전트가 수행할 수 있는 행동의 집합으로 정의된다. 배전계통 재구성 문제에서 행동은 목적함수의 의사결정 변수인 개폐기의 개/폐 상태로 정의된다. 본 논문에서 개폐 상태는 시간 t에서 개폐기 l이 개방된 상태를 0, 단락된 상태를 1로 하여 식(10)과 같이 sw_l^t로 정의하였고, 절체용 개폐기의 수를 L이라 할 때 시간 t에서 에이전트의 행동 a^t를 식(11)와 같이 정의하였다.

$$sw_l^t = \begin{cases} 0 & \text{if } switch\ off \\ 1 & \text{if } switch\ on \end{cases} \tag{10}$$

$$a^t = (sw_1^t,\ sw_2^t,\ \cdots,\ sw_L^t) \tag{11}$$

4.2 상태(state)

상태란 에이전트가 활동하는 환경에 대한 상태를 의미한다. 강화학습에서는 에이전트의 행동에 따라 상태가 새로운 상태로 이전된다. 본 논문에서는 상태를 시간 t와 기준모선을 제

외한 모선 n의 유효전력 p_n^t을 성분으로 하는 벡터 $\boldsymbol{p}$, 유효전력 q_n^t을 성분으로 하는 벡터 $\boldsymbol{q}$, 시간 $t-1$의 행동 a^{t-1}으로 정의하였다. 상태는 식 (12)와 같이 표현된다.

$$s^t = (t,\ \boldsymbol{p},\ \boldsymbol{q},\ a^{t-1}) \tag{12}$$

4.3 보상(reward)

보상은 주어진 상태에서 취해진 행동에 대한 결과로 에이전트가 획득하는 이득이다. 보상은 상태와 행동에 대한 함수로 표현된다. 합리적인 에이전트는 운영 기간이 종료되었을 때까지 받는 보상의 합인 누적보상을 최대화하는 방향으로 학습하기 때문에 보상의 설정은 학습효과 및 결과에 지대한 영향을 미친다. 본 논문에서는 보상을 최적화 문제에서 정의한 전력손실인 출력삭감(c_n^t), 선로손실(l^t) 및 개폐기 동작에 의한 개폐 손실(sl^t)의 합에 음의 부호를 취하여 식 (13)과 같이 정의하였다. 보상을 구성하는 출력삭감 및 선로손실은 조류계산을 통해 구할 수 있다.

$$r(s^t,a^t) = -((c_n^t + l^t)\ \bullet\ \triangle t + sl^t) \tag{13}$$

4.4 DQN 학습과정

DQN의 전반적인 학습과정인 표 1을 정리하면 다음과 같다. 알고리즘에서 사용된 D는 리플레이 버퍼이고, M 은 에피소드의 수, T는 총 시간, C는 Q-목표 파라미터 갱신 주기, γ는 감가율을 의미한다. 알고리즘에 대한 자세한 설명은 다음과 같다. (1-2) 리플레이 버퍼 D와 행동 가치함수 파라미터 θ^Q을 초기화하고, 효율적인 학습을 위해 식 (12)로 구성된 상태행렬의 정규화 과정을 거친다. (3-4) 시간 1에서 T로 구성된 에피소드에 대해 (4-14)를 M번 반복한다. (5-8) DQN에 현재 상태를 입력으로 넣어주면, 해당 상태에서 가능한 모든 Q값이 반환되고 에어전트는 ε-greedy 정책에 따라 ε의 확률로 무작위 행동을 하고, 1-ε의 확률로 최대 Q값을 갖는 행동을 수행한다. 본 논문에서는 학습 초기에는 에이전트가 모델에 대해 충분히 탐색하고 후반으로 갈수록 학습한 행동을 활용하기 위해 k를 두어 ε값을 점차적으로 줄여갔다. (9) 에이전트가 행동(개폐동작)을 취하면 그에 따른 보상을 취득하고 다음 상태로 이동한다. 상태, 행동, 보상, 다음 상태의 일련의 과정인 경험을 튜플로 구성하여 리플레이 버퍼에 저장한다. (10) 버퍼에 충분한 양의 경험이 쌓이면 버퍼에서 배치크기 만큼의 경험을 무작위로 선택하여 DQN을 학습시키는 데 사용한다. 리플레이 버퍼에서 경험튜플을 임의로 선택하면 경험 간의 상관성이 줄어들어 에이전트가 특정 경험에 과적합되지 않고 다양한 경험을 통해 학습 효율을 높일 수 있다. (11-12) 추출된 경험으로 손실을 계산하고 손실을 최소화하기 위해 Q 신경망의 파라미터 θ^Q에 대한 경사강하법을 수행한다. (13-14) C번째 에피소드마다 Q-목표 신경망의θ^{Q-}을 Q 신경망의 θ^Q로 갱신한다. (16) 지정된 에피소드가 끝나면 DQN의 가중치와 편향을 저장한다.

표 1 DQN 학습 알고리즘

Table 1 DQN learning algorithm

DQN 학습 알고리즘
1 : 리플레이 버퍼 D, $Q(s,\ a;\theta^Q)$의 파라미터 θ^Q 초기화
2 : 파라미터: $0<\gamma<1,\ 0<\epsilon_{\max}<1,\ 0<\epsilon_{\min}<1,\ 0<k<1$
3 : for $i = 1,...,M$
4 : for $t = 1,...,T$
5 : $\epsilon = \max(\epsilon_{\max} - k \bullet i, \epsilon_{\min})$
6 : ϵ의 확률로 무작위 행동 a_t 수행
7 : $1-\epsilon$의 확률로 $a^t = \arg\max_{a'} Q(s,a';\theta^Q)$ 수행
8 : 해당 보상 값인 r^t 취득 후 새로운 상태인 s^{t+1}로 이동
9 : 리플레이 버퍼 D에 튜플 (s^t,a^t,r^t,s^{t+1}) 저장
10 : 리플레이 버퍼 D에서 배치 크기 만큼 무작위 추출
11 : θ^Q에 관하여 $L(\theta^Q)$에 경사강하법 수행
12 : $\min_{\theta} \sum_{t=0}^{T} [r^t + \gamma \max_{a'} Q(s^{t+1},a';\theta^{Q-}) - Q(s^t,a^t;\theta^Q)]^2$ endfor
13 : *If* $\mathrm{mod}(i,C) = 0$
14 : Q -목표 신경망 파라미터 업데이트 ($\theta^{Q-} \leftarrow \theta^Q$)
15 : endfor
16 : $Q(s,a;\theta)$ 출력

4.5 개폐 동작 제한을 위한 알고리즘 보정

에이전트는 강화학습의 목적인 누적보상의 최대화를 위해서 현재의 행동이 손해가 나더라도 최종적으로 더 높은 누적보상을 가져온다면 그 행동을 취한다. 그러나 학습을 통해 구한 보상 값은 모든 경우를 학습하는 것은 아니기에 실제 동작 시에 얻게 되는 보상 값과 오차가 존재한다. 예를 들어 누적보상이 더 높아 현재 손해를 보는 행동을 수행했는데 학습 데이터와 검증 데이터간의 보상 값의 오차로 인하여 검증 시에는 누적보상이 더 낮은 현재 손해를 복구하지 못하는 경우가 발생할 수 있다. 이와 같은 경우에는 행동을 취하지 않고 현재 토폴로지를 유지하는 것이 더 좋은 결과가 된다. 위와 같은 문제를 해결하기 위해 본 논문에서는 에이전트의 의사결정 과정의 마지막 단계에서 식 (14)의 알고리즘을 추가하여 학습 오차로 야기된 행동의 오동작을 줄일 수 있도록 하였다. 식 (14)는 임계값(λ)을 두어 DQN에서 도출된 행동에 대한 보상과 개폐 동작을 하지 않았을 때의 보상의 차이가 임계값을 넘는 경우에만 해당 개폐를 수행하는 알고리즘이다. 즉, 제안하는 보정 알고리즘은 누적보상의 오차로 발생하는 손해의 임계값을 설정함으로써 현재 큰 손해를 감내한 행동을 무시하여 개폐기의 불필요한 개폐 동작을 방지한다.

$$a^t = \begin{cases} a^{t-1} & \text{if } r(t, \arg\max_{a'} Q(s, a';\theta)) - r(t, a^{t-1}) \le \lambda \\ \arg\max_{a'} Q(s, a';\theta) & \text{if } r(t, \arg\max_{a'} Q(s, a';\theta)) - r(t, a^{t-1}) > \lambda \end{cases} \quad (14)$$

5. 사례 연구

제안하는 강화학습 기반 배전계통 재구성 기법의 성능을 평가하기 위해 사례 연구를 진행하였다. 사례 연구 결과에서는 제안하는 기법을 개폐를 전혀 하지 않는 고정기법, 무작위 개폐 동작기법, 시간별 최적화기법과의 전력손실을 비교하였다.

5.1 사례 연구 구성 및 강화학습 설정

본 사례 연구에 사용된 모델 계통의 용량은 한국의 특별고압(22.9 kV) 배전망 선로를 기준으로 15 MVA로 설정하였고 배전망 토폴로지는 [24]의 배전계통 33 테스트 모선을 차용하였다. 각 모선의 부하데이터는 미국 중서부 지역의 2017년 실제 모선의 스마트미터 전력 데이터를 활용하였는데, 특별고압 일반 배전선로의 상시운전용량이 10 MVA 인 점을 고려하여 전력 데이터를 상수배하여 최대값이 약 10 MVA가 되게 조정하였다.[25] 각 부하모선의 역률은 0.9로 일정하다고 가정하였다. 전압 유지범위는 국내 특별고압 일반 배전망의 기준을 따라 0.91 p.u.이상, 1.04 p.u.이하로 설정하고 허용전류 최댓값은 ACSR-OC 160mm^2 선로를 기준으로 395 A로 설정하였다. 절체용 개폐기 5개는 [24]의 테스트 모선과 동일한 구성으로 그림 2와 같이 (33), (34), (35), (36), (37) 위치에 있다고 가정하였다. 개폐기의 개폐 손실은 [20]의 개폐 동작비용의 산정방식을 참조하여 동작 비용을 산출하고 이를 전력량으로 환산한 값인 5 kWh로 설정하였다. 일반적인 배전선로는 최대 12 MW까지 분산전원을 접속할 수 있도록 규정하고 있는 점을 고려하여 태양광 발전기의 용량을 12 MW이하에서 출력삭감이 일어나는 용량인 8.5 MW로 상정하였다.[26] 태양광 발전기는 그림 2의 모선 18번에 설치되었다고 가정하였다.

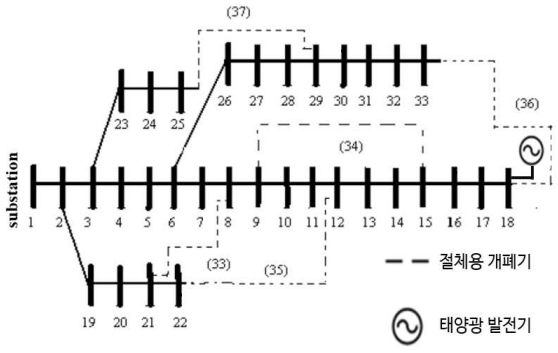


그림 2 배전계통 33 테스트 모선
Fig. 2 A 33-bus radial distribution system

주어진 부하와 발전량으로 시간 단위의 조류계산을 수행하여 토폴로지별 제한전력과 선로손실을 산출하고 강화학습의 학습에서 주어지는 보상을 산정하였다. Q 신경망을 학습 데이터로서 25일(2017년 1월 1일에서 2017년 1월 25일)의 전력 데이터를 사용하였고, 검증 데이터로 10일(2017년 1월 26일에서 2017년 2월 4일)의 전력데이터를 사용하였다. 본 연구에서 사용한 DQN의 주요 파라미터 정보는 표 2와 같다.

표 2 DQN 하이퍼파라미터
Table 2 Hyperparameters of DQN

하이퍼파라미터	값
학습율(α)	0.0005
감가율(γ)	0.98
리플레이 버퍼 크기(D)	50000
배치 크기	128
초기 입실론(ϵ_o)	0.2
Q-목표 업데이트 주기(C)	60 에피소드
뉴런 수	입력층: 38, 은닉층(2): 600, 출력층: 32

5.2 결과 분석

제안하는 배전계통 재구성 기법의 학습과정을 그림 3에 도시하였다. 그림 3의 결과와 같이 학습 데이터에 대한 총 전력손실은 에피소드가 증가함에 따라 점차적으로 감소하였고, 에피소드가 약 70이상에서 수렴하는 것을 확인하였다. 모의실험에 사용된 컴퓨터의 사양은 Intel Xeon CPU, NVIDIA GeForce GTX 1080, 64 GB이다. 한 에피소드를 학습하는데 약 1분 30초의 시간이 소요되어 DQN이 수렴하는데 총 2시간 정도가 소요되었다.

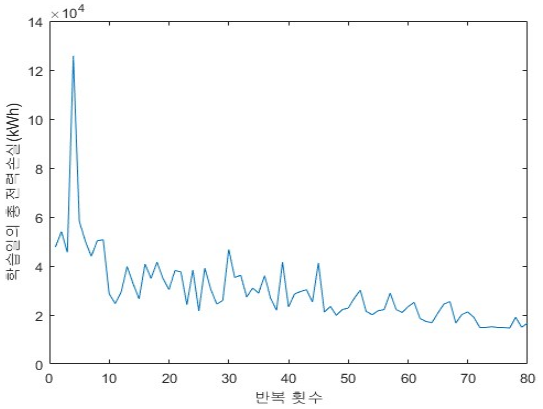


그림 3 DQN 학습과정
Fig. 3 DQN training process

표 3에서는 기법별로 검증 데이터의 일평균 전력손실과 개폐 동작횟수를 확인했다. 일평균 전력손실은 검증일의 전체 손실을 전체 일수로 나눈 값으로 정의하였다. 배전계통 재구성 없는 고정 토폴로지의 경우, 선로손실과 출력삭감은 각각 462 kWh, 6235 kWh이였다. 전력손실의 대부분을 차지하는 출력삭감은 해당 계절인 겨울철 태양광 발전기의 하루 발전량인 23154 kWh의 약 27%에 해당하는 큰 비효율을 야기한다. 실험 결과 무작위 개폐 동작만으로도 전력손실이 토폴로지를 고정했을 때에 비해 크게 감소하는 것으로 확인되었다. 그 외에도 시간별 최적화 기법, 동적 계획법, DQN 기법, 보정된 DQN 기법을 배전계통 재구성 문제에 적용하여 결과를 비교하였다. 이들 기법은 토폴로지를 고정했을 때의 경우에 비해 약 8%에 해당하는 전력손실이 발생하였다. 즉, 배전계통재구성은 전력손실 측면에서 큰 효과를 내는 것을 알 수 있었다. 최적화 기법인 동적계획법은 상태전이 함수를 알 때 사용하는 기법으로 과거의 제어주기 동안의 이상적인 해를 도출할 수 있지만 현실에서는 예측데이터에 대한 이상적인 해를 이용하기 때문에 제어성능이 예측정확도에 의존하는 특징이 있다. 시간별 최적화 기법은 모든 토폴로지에 대해서 실시간 조류계산을 진행하여 보상 정보를 안다는 가정 하에 최적의 토폴로지를 선택하는 기법으로 매시간 조류계산을 해야 한다는 단점이 있다. DQN 기법은 상태전이 함수를 모르는 상태에서 데이터를 통해 스스로 학습하면서 동작한다는 장점을 가지고 있으며 시간별 최적화와 유사한 성능을 보이는 것을 확인하였다.

표 3 기법별 평균 전력손실 및 개폐 동작횟수 (2017년 겨울)
Table 3 Average power loss and switching number (winter 2017)

기법	평균 전력손실 (kWh)	평균 개폐 동작횟수
토폴로지 고정	6697	0
무작위 개폐	2476	61.7
시간별 최적화	576	7.2
동적 계획법	532	6.2
DQN	597	11.2
DQN(보정)	561	8.9

그러나 DQN 기법의 개폐 동작횟수는 일평균 11.2회로 다소 빈번한 개폐 동작을 보였으며 그로 인하여 전력손실도 시간별 최적화와 동적 계획법에 못 미치는 결과가 나왔다. 이러한 DQN의 한계를 극복하기 위하여 본 논문에서는 임계값 기반의 단순한 판단을 최후에 추가하였다. 적절한 임계값을 산정하기 위하여 [그림 4]와 같이 임계값을 -30 kWh에서 5 kWh씩 증가시키면서 검증일의 평균 전력손실과 평균 개폐 동작횟수를 분석하였다. 임계값이 증가함에 따라 동작횟수는 감소하였고 전력손실은 감소하다가 증가하는 추이를 보였다. 이는 DQN의 학습 데이터와 검증 데이터 간의 오차에 의해 불필요한 개폐 동작을 하여 전력손실을 발생시킨다는 것을 의미한다. 따라서 보정단계에서의 임계값은 검증일의 평균 전력손실을 최소화하는 -10 kWh로 설정하였다. 보정단계를 적용한 DQN기법은 기존 DQN 기법 대비 전력손실은 596 kWh에서 561 kWh로, 개폐 동작횟수는 11.2회에서 8.9회로 감소하였다. 이 결과는 시간별 최적화보다 15 kWh 적은 전력 손실 성능을 보이는 것이다.

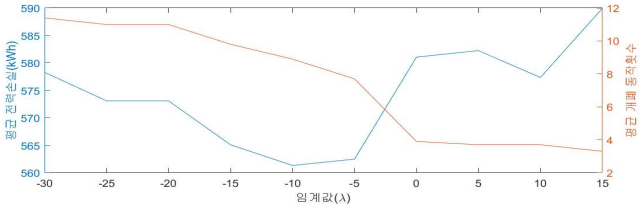


그림 4 임계값에 따른 전력손실 및 개폐 동작횟수
Fig. 4 Power loss and switching number according to the threshold

표 3에서 비슷한 성능을 보였던 동적 계획법, 시간별 최적화, DQN, 보정된 DQN에 대한 결과 중 임의의 날에 대해서 그림 5의 그래프와 같이 전력손실을 시간 단위로 가시화하였다. 표 4는 그림 5에 해당하는 시간별 토폴로지의 변화를 의미하며 토폴로지는 식 (11)과 같이 $(sw_{33}^t, sw_{34}^t, sw_{35}^t, sw_{36}^t, sw_{37}^t)$로 구성된다. DQN의 경우 동적 계획법과 유사하게 처음부터 개폐 동작을 하여 뒤에 오는 선로손실에 대비하였고 오전 8시에 추가로 개폐 동작을 하여 낮 시간에 발생하는 출력삭감을 사전에 대비하는 것을 확인할 수 있었다. 이를 통해 에이전트는 즉각적인 보상에만 의존하지 않고 뒤에 오는 보상까지 고려한 누적보상을 극대화시키는 행동을 취하는 것을 확인하였다. 그러나 불필요하게 많은 개폐 동작을 하는 것을 개선하기 위하여 제안한 보정된 DQN 기법은 시간별 최적화 기법과 유사하게 동작하는 것을 확인하였다. 그러나 보정된 DQN 기법 역시 누적보상을 고려하기에 시간별 최적화 기법보다 높은 성능을 보인다.

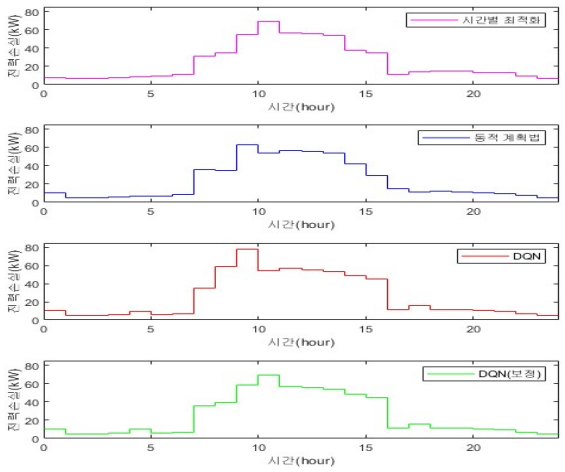


그림 5 기법별 검증 데이터 결과 비교
Fig. 5 Comparison of Simulation results

표 4 시간별 토폴로지 변화 (sw_{33}^t, sw_{34}^t, sw_{35}^t, sw_{36}^t, sw_{37}^t)

Table 4 Topology change over time

시간	시간별 최적화	동적 계획법	DQN	DQN보정
0시	(0,0,0,0,0)	(0,0,1,0,0)	(0,0,1,0,0)	(0,0,1,0,0)
4시	(0,0,0,0,0)	(0,0,1,0,0)	(0,0,1,0,1)	(0,0,1,0,1)
7시	(0,0,0,0,0)	(0,0,0,0,0)	(0,0,1,0,1)	(0,0,1,0,1)
8시	(0,0,0,0,0)	(0,0,0,0,0)	(1,1,0,0,1)	(0,0,1,0,1)
9시	(1,1,0,0,0)	(0,0,0,1,0)	(0,0,0,1,0)	(0,1,1,0,0)
10시	(0,0,0,1,0)	(0,0,0,1,0)	(0,0,0,1,0)	(0,0,0,1,0)
14시	(0,0,0,0,0)	(0,0,0,0,1)	(0,1,0,0,0)	(0,1,0,0,0)
15시	(0,0,0,0,1)	(0,0,0,0,1)	(0,1,0,0,1)	(0,1,0,0,1)
16시	(0,0,0,0,1)	(0,0,1,0,1)	(0,1,0,0,1)	(0,1,0,0,1)
17시	(0,0,0,0,1)	(0,0,1,0,1)	(0,1,1,0,1)	(0,1,1,0,1)

겨울 이외의 다른 계절에서도 모의실험을 진행하였다. 표 5는 2017년 5월 1일에서 2017년 5월 10일의 평균 전력손실 및 개폐 동작횟수이다. 봄철에 해당하는 5월의 평균 전력손실은 겨울철인 1월과 2월에 비해 토폴로지를 고정한 경우 약 4배가 증가되는 것을 확인할 수 있었다. 이는 태양광 발전량의 증가로 인한 출력삭감이 증가하였기 때문이다. 5월의 경우, DQN 기반 배전계통 재구성을 적용하였을 때의 평균 전력손실이 토폴리지를 고정했을 때의 약 3%로 겨울철에 비해 큰 감소율을 보였다. 가을의 경우 봄과 유사한 결과를 보였으나 여름의 경우에는 DQN의 성능이 다른 계절에 비하여 저하되는 것을 확인하였다. 여름철에 성능이 저하되는 이유는 다양한 토폴로지에서 출력삭감이 발생하기에 이전보다 더 많은 학습이 필요하기 때문이다. 여름철 특징을 보다 잘 학습할 수 있는 기법에 대한 후속 연구가 필요하다.

표 5 기법별 평균 전력손실 및 개폐 동작횟수 (2017년 봄)

Table 5 Average power loss and switching number (spring 2017)

기법	평균 전력손실 (kWh)	평균 개폐 동작횟수
토폴로지 고정	27612	0
무작위 개폐	12728	61.2
시간별 최적화	826	11.9
동적 계획법	812	11.4
DQN	847	13
DQN(보정)	839	9.6

6. 결 론

본 논문에서는 재생에너지가 다수 설치된 배전망에서 선로손실 및 재생에너지 활용을 극대화하는 강화학습기반 배전계통 재구성 기법을 제안하였다. 강화학습의 기법중 하나인 Q-러닝은 모델에 대한 정확한 정보 없이 스스로 학습하여 행동하기 때문에 다양한 배전계통에 효과적으로 활용할 수 있다. 하지만 강화학습을 활용한 배전계통 재구성 기법은 경우의 수가 많은 상태에 대한 판단에 오차가 발생하여 과다한 개폐 동작과 전력손실을 발생시키는 한계점을 보였다. 따라서 이러한 문제점을 보완하기 위하여 최종 보정과정을 두어 강화학습이 결정한 행동에 대해 제한을 두는 기법을 제안하였다. 배전계통 33 테스트 모선을 사용한 사례 연구를 통해 보정과정을 포함한 강화학습의 효용성을 입증하였다. 제안하는 기법은 전역 데이터 없이 과거 데이터를 통한 학습으로도 모든 정보를 알고 동작하는 최적화 기법과 거의 유사한 성능을 보이는 것을 확인하였다. 향후 연구에서는 다수의 태양광 발전기를 이용하여 전압 유지범위 내에서 최대로 유입될 수 있는 제한전력의 집합을 구성하여 발전 사업자의 전체 이익을 최대화하는 방식을 고려할 예정이다. 추가로 여름철에도 원활히 학습할 수 있는 DQN 구조에 대한 연구가 필요하다.

Acknowledgements

This work was supported in part by the “Human Resources Program in Energy Technology” initiative of the Korea Institute of Energy Technology Evaluation and Planning (KETEP), granted financial resource from the Ministry of Trade, Industry & Energy, the Republic of Korea (No. 20184010201690), and in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2020R1F1A1075137).

References

[1] Ministry of Trade, Industry and Energy, the Republic of Korea, “재생에너지 3020 이행계획(안) 발표,” https://url.kr/KQLCqB, 2017.

[2] Ministry of Trade, Industry and Energy, the Republic of Korea, “1 MW 이하 소규모 신재생발전 전력망 접속보장,” https://url.kr/8phdEH, 2016.

[3] M. M. Haque and Peter Wolfs. “A review of high PV penetrations in LV distribution networks: Present status, impacts and mitigation measures,” Renewable and Sustainable Energy Reviews 62, pp. 1195-1208, 2016.

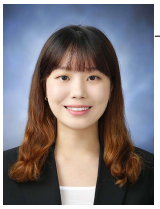
[4] J. Seuss, M. J. Reno, R. J. Broderick, and S. Grijalva, “Improving distribution network PV hosting capacity via smart inverter reactive power support,” 2015 IEEE Power & Energy Society General Meeting, pp. 1-5, 2015.

[5] Y. Kim, H. Myung, N. Kang, C. Lee, M. Kim and S. Kim, “Operation Plan of ESS for Increase of Acceptable Product of Renewable Energy to Power System,” The Transaction of KIEE 67.11, pp. 1401-1407, 2018.

[6] F. Capitanescu, L. F. Ochoa, H. Margossian, and N. D. Hatziargyriou, “Assessing the potential of network reconfiguration to improve distributed generation hosting capacity in active distribution systems,” IEEE Transactions on Power

Systems 30.1, pp. 346-356, 2014.
[7] H. Myung and S. Kim "The Study on the Method of Distribution of output according to Power Limit of Renewable Energy," The Transaction of KIEE 23.1, pp. 173-180, 2018.
[8] A. G. Patel and C. Patel, "Distribution network reconfiguration for loss reduction," International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), IEEE, pp. 3937-3941, 2016.
[9] A. M. Imran and M. Kowsalya, "A new power system reconfiguration scheme for power loss minimization and voltage profile enhancement using fireworks algorithm," International Journal of Electrical Power & Energy Systems 62, pp. 312-322, 2014.
[10] B. Novoselnik and M. Baotić, "Dynamic reconfiguration of electrical power distribution systems with distributed generation and storage," IFAC-PapersOnLine, vol. 48, no. 23, pp. 136-141, 2015
[11] E. A. Feinberg, J. Hu, and K. Huang, "A rolling horizon approach to distribution feeder reconfiguration with switching costs," In: 2011 IEEE International Conference on Smart Grid Communications (SmartGridComm), IEEE, pp. 339-344, 2011.
[12] F. V. Dantas, D. Z. Fitiwi, S. F. Santos and J. P. S. Catalao, "Dynamic reconfiguration of distribution network systems: A key flexibility option for res integration," in 2017 IEEE International Conference on Environment and Electrical Engineering and 2017 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe), pp. 1-6, June 2017.
[13] M. Mosbah, S. Arif, R. D. Mohammedi and A. Hellal, "Optimum dynamic distribution network reconfiguration using minimum spanning tree algorithm," in 2017 5th International Conference on Electrical Engineering - Boumerdes (ICEE-B), pp. 1-6, Oct 2017.
[14] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez and Y. Chen, "Mastering the Game of Go without Human Knowledge," Nature - International Journal of Science, vol. 550, pp. 354-359, Oct 2017.
[15] M. Bellemare, Y. Naddaf, J. Veness, and M. Bowling, "The arcade learning environment: An evaluation platform for general agents," Journal of Artificial Intelligence Research, vol. 47, pp. 253-279, 2013.
[16] J. Kober, J. Bagnell and J. Peters, "Reinforcement learning in robotics: a survey," International Journal of Robotics Research, vol. 32(11), pp. 1238-1278, 2013.
[17] S. Kim and H. Lim, "Reinforcement learning based energy management algorithm for smart energy buildings," Energies, vol. 11, no. 8, pp. 2010, 2018.
[18] T. Li, Y. Xiao and L. Song, (2019, October), "Deep Reinforcement Learning Based Residential Demand Side Management With Edge Computing," In 2019 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm) IEEE, pp. 1-6, 2019.
[19] Q. Yang, G. Wang, A. Sadeghi, G. B. Giannakis and J. Sun, "Two-Timescale Voltage Control in Distribution Grids Using Deep Reinforcement Learning," in IEEE Transactions on Smart Grid, vol. 11, no. 3, pp. 2313-2323, May 2020.
[20] Y. Gao, J. Shi, W. Wang and N. Yu, "Dynamic Distribution Network Reconfiguration Using Reinforcement Learning," 2019 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm), Beijing, China, pp. 1-7, 2019.
[21] S. S. Gu, T. Lillicrap, R. E. Turner, Z. Ghahramani, B. Schölkopf and S. Levine, "Interpolated policy gradient: Merging on-policy and off-policy gradient estimation for deep reinforcement learning," Advances in neural information processing systems, pp. 3846-3855, 2017.
[22] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra and M. Riedmiller, "Playing atari with deep reinforcement learning," arXiv preprint arXiv: 1312.5602, 2013.
[23] T. Markvart, A. McEvoy and L. Castaner, "Practical handbook of photovoltaics: fundamentals and applications," Elsevier, pp. 49-51, 2003.
[24] J. Z. Zhu, "Optimal reconfiguration of electrical distribution network using the refined genetic algorithm," Electric Power Systems Research 62.1, pp. 37-42, 2002.
[25] Iowa State University, "A Real 240-Node Distribution System with One-Year Smart Meter Data," http://wzy.ece.iastate.edu/Testsystem.html, 2017.
[26] S. Kim, "Increasing Hosting Capacity of Distribution Feeders by Analysis of Generation and Consumption," KEPCO Journal on Electric Power and Energy Vol. 5, No. 4, December 2019, pp. 295-309, 2019.

저자소개



Se-Heon Lim

She received her B.S. degree in Electrical Engineering from Soongsil University, Seoul, South Korea, in 2018. Currently, she is pursuing Ph.D. degree at Soongsil University, Seoul, Korea.
E-mail: seheon0223@naver.com

Tae-Geun Kim

He received his B.S degree in from Department of electrical and electronics engineering from Kangwon University, Chuncheon, South Korea, in 2020. Currently, he is pursuing M.E. degree at Soongsil University, Seoul, Korea.
E-mail: taegeun1520@gmail.com

Sung-Guk Yoon

He received the B.S. and Ph.D. degrees in electrical engineering and computer science from Seoul National University, Seoul, South Korea, in 2006 and 2012, respectively. He is currently with Soongsil University as an associate professor. His research include energy big data, game theory for power system, and power system optimization.

E-mail: sgyoon@ssu.ac.kr