

최소가동시간 제약을 고려한 심층 강화학습 기반의 다중 냉동기 운영 모델

A Multi-chiller Operation Model Based on Deep Reinforcement Learning Considering Minimum Up-time Constraint

김종은¹ · 김관호² · 김재곤[†]

인천대학교 산업경영공학과¹, 동국대학교 산업시스템공학과²

요약

여름철 냉동기가 건물의 주 에너지 소비자로 고려됨에 따라 효율적인 냉동기 운영은 매우 중요한 문제로 고려된다. 그러나, 건물의 냉방수요가 건물 내외부 환경, 건물 재실자의 행동 등의 여러 요인에 의해 변동하고 냉동기의 가동제약조건으로 인해 현재 시점의 운영이 미래 시점의 운영에 제약을 발생시킴에 따라 건물의 냉방수요를 정확하게 만족하도록 냉동기를 운영하는 것은 어렵다. 본 연구에서는 이러한 문제를 해결하기 위해 냉동기의 최소화동시간을 고려한 심층 강화학습 기반의 다중 냉동기 운영 모델을 제안한다. 제안한 모델은 외기 정보와 냉방시스템 내부 정보로 구성된 상태에 따른 냉동기 운영 조합이 갖는 가치를 학습하고 실현 가능한 냉동기 운영 중 건물의 냉방수요와 냉동기에 의한 공급 부하 간의 차이를 최소화할 수 있는 냉동기 운영 조합을 결정한다. 냉동기의 최소화동시간 제약을 고려한 훈련 알고리즘을 적용하여 제안한 모델의 현실 적용 가능성을 높였으며 실제 국내 A대학교의 데이터를 바탕으로 실험한 결과, 제안한 다중 냉동기 운영 모델이 최소화동시간을 준수함과 동시에 건물 냉방 부하와의 차이 측면에서 A대학교의 기존 냉동기 운영 로직보다 우수한 성능을 보임을 확인하였다.

■ 중심어 : 강화학습, 냉동기 운영, 중앙 냉방시스템, 건물 에너지 관리 시스템

Abstract

In summer, as chillers are considered the main energy consumer of building, the efficient chiller operation is considered important. However, it is difficult to operate chillers to meet the cooling demand of the building as the demand fluctuates with various factors like the internal, external environment and behavior of the occupants and as chiller's constraint cause the current operation constrains operation in future. To address these problems, this study proposes a multi-chiller operation model based on deep reinforcement learning considering the minimum up-time of the chiller. The proposed model learns the value of the chiller operations according to the state composed of metrological and cooling system information and determines operation that minimizes the difference between the supply load and the cooling demand among feasible operations. The practical applicability was improved by applying the training algorithm considering the minimum up-time constraint and Experiments results using the actual data from a Korean university confirmed that the proposed model complies with the chiller constraints and outperforms the existing chiller operation logic of the university in terms of differences from the building cooling demand.

■ Keyword : Reinforcement Learning, Chiller Operation, Central Cooling System, Building Energy Management System(BEMS)

2024년 05월 30일 접수; 2024년 06월 10일 수정본 접수; 2024년 06월 12일 게재 확정.

* 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 학석사연계ICT핵심인재양성사업의 연구결과로 수행되었음 (IITP-2024-RS-2023-00260175)

† 교신저자 (jaegkim@inu.ac.kr)

I. 서론

건물 냉방시스템 에너지 사용량의 40% 이상이 냉동기 운영을 위해 소비됨에 따라, 에너지 절감을 위한 냉동기 운영은 중요한 요소로 고려된다[15]. 특히, 대학 캠퍼스와 같이 여러 건물에 냉방에너지를 공급하기 위해 사용되는 중앙 냉방시스템은 다수의 냉동기를 운영하는 만큼 효율적인 냉동기 운영 제어가 필수적이다. 냉방시스템 관리자는 서로 다른 정격 용량을 갖는 냉동기들의 ON/OFF 제어를 통해 각 건물에 공급할 냉방부하를 결정한다. 건물의 냉방수요를 초과하는 냉방부하 공급은 에너지 낭비와 추가적인 비용을 발생시키는 반면 수요에 비해 부족한 공급은 건물 내 쾌적한 환경을 저하시킬 수 있다. 따라서 냉동기는 건물의 냉방부하 수요와 가장 근접하게 운영되어야 하며 최적 냉동기 운영을 통해 불필요한 냉동기 가동을 줄여 에너지 손실을 방지함과 동시에 재실자의 쾌적한 실내 환경을 유지하며 냉동기 운영 효율을 높일 수 있다[23].

그러나 건물의 냉방부하 수요에 따른 최적 냉동기 운영은 외기 환경, 냉방시스템의 상태, 냉동기의 최소 가동 시간과 같은 다양한 요소에 대한 고려가 필요하기 때문에 매우 어려운 문제로 고려된다[8, 22]. 건물의 냉방부하는 온도, 습도, 재실자의 행동 패턴과 같은 요인들의 영향을 받아 변화하는 값으로 정확한 냉방부하를 냉동기 운영에 반영하기 어렵다. 또한 냉동기를 가동함에 있어 고려해야 할 최소가동시간이 존재한다[2]. 냉동기의 수명 유지를 위해 각 냉동기는 가동 시 최소가동시간 동안 가동 상태를 유지한 후 정지되어야 하며 최소가동시간 이전에 정지될 수 없다. 이러한 냉동기의 가동 조건으로 인해 현재 시점과 더불어 미래 시점의 변화를 고려한 냉동기 운영이 요구되며 문제의 복잡성이 높다.

이러한 어려움에도 불구하고 효율적인 냉동기 운영을 위한 연구는 활발히 진행되었다. 건물의 냉방부하에 따른 최적 냉동기 제어 전략을 도출하기 위해 동적 프로그래밍(Dynamic Programming), 입자 군집 최적화(Particle Swarm Optimization), 뻐꾸기 탐색(Cuckoo Search Algorithm)과 같은 최적화 알고리즘이 사용되었다[1,2,3]. 그러나 최적화 알고리즘의 경우, 건물의 냉방부하에 대한 가정이나 예측 값에 대한 의존도가 높고 실제 냉방시스템 운영에서 발생하는 다양한 변동을 반영하지 못한다[16].

이에 따라, 실제 발생하는 변동에 의한 냉동기 운영 오류를 줄이기 위해 예측 모델을 결합한 냉동기 운영 모델들이 제안되었다. Yanzhu et al.[22]은 혼합 정수 선형계획법 기반으로 미래 24시간 예측 값에 대한 최적 운영을 수립하고 휴리스틱 룰을 추가로 구축하였다. Sala-Cardoso et al.[19]은 순환신경망 기반의 건물의 냉방부하 예측 모델과 수리모델링을 결합하였으며 Liao et al.[13]은 자기회귀 이동평균 기반의 교정 모델을 통해 실시간 변동에 대한 안정성을 높였다. 그러나 기존 연구에서 제안한 모델들 역시 예측 모델의 성능이 냉동기 운영 모델의 성능에 직접적인 영향을 미치며 냉동기의 최소가동시간을 고려하지 않아 본 연구에서 고려하는 문제에 적용하기 어렵다.

최적화 분야에서 우수한 성능을 보이는 강화 학습을 적용한 냉동기 운영 연구도 진행되었다[11]. Du et al.[4]은 Deep Q-Network(DQN) 기반의 건물 냉방시스템 제어 모델을 제안하였으며, 냉동기의 공급온도를 조절함으로써 단일 건물 환경에서 에너지 사용량을 절감시켰다. Wong et al.[20]은 계층적 강화 학습 기반의 냉동기 제어 모델을 제안하였으며 제안 모델은 계층 구조를 통해 정해진 실내 온도 범위를 만족하는 냉동기의 제어 전략을 수립하였다. He et al.[5]은 장단기 기억 신경망(Long Short Term Memory, LSTM)

과 DQN을 결합하여 냉수의 공급온도를 제어하였으며 시뮬레이션 환경에서 기존의 규칙 기반의 제어보다 에너지 절약 효과가 뛰어남을 확인하였다.

냉동기 운영 효율을 높이기 위한 기존 연구들은 주로 단일 건물의 냉방시스템을 대상으로 진행되었다[13,16,19]. 그러나 다수의 건물에 냉방을 공급하는 중앙 냉방시스템은 단일 건물의 냉방시스템보다 큰 규모와 복잡한 구조를 가진다. 냉수가 다수의 건물 사이를 순환함에 있어 발생하는 변동과 같이 단일 건물의 운영 모델은 중앙 냉방시스템의 특성을 충분히 반영할 수 없다. 또한 냉동기 운영 전략 수립에 있어 냉동기 수명 유지를 위한 최소가동시간을 고려한 연구가 현저히 적다. 최소가동시간은 냉방시스템의 유지보수 비용을 줄이기 위해 고려되어야 하며 이에 따라 운영 시간 제약을 만족시키며 에너지를 절감시킬 수 있는 냉동기 운영 모델이 요구된다.

따라서 본 연구에서는 중앙 냉방시스템에서 최소가동시간을 고려한 심층 강화학습 기반의 다중 냉동기 운영 기법을 제안한다. 동적으로 변화하는 냉방시스템 환경에 대응하기 위해, 환경 변화에 따라 최적의 행동을 결정하는 DQN 기반의 다중 냉동기 운영 모델을 구축한다. 제안한 모델은 외기 환경 정보와 냉방시스템 내부 정보로 구성된 상태 값을 바탕으로 최소가동시간 제약을 준수하며 건물의 냉방부하와 가장 근접한 냉동기 운영 조합을 결정하도록 학습한다. 냉동기 운영에 의한 냉방시스템의 내부 환경 변화를 반영하기 위해 실제 데이터를 활용하여 실제 냉방시스템 환경을 모사한 시뮬레이터를 구축하였으며 이러한 시뮬레이터는 냉동기 운영에 따른 냉방시스템의 환경 변화를 예측하고 제안한 모델의 성능을 검증하는데 사용된다.

구체적으로 다중 냉동기 운영 모델의 상태는 온도, 습도와 같은 외기 환경 정보와 냉수의 공급, 환수온도 및 냉동기 운영 상태와 같은 냉방

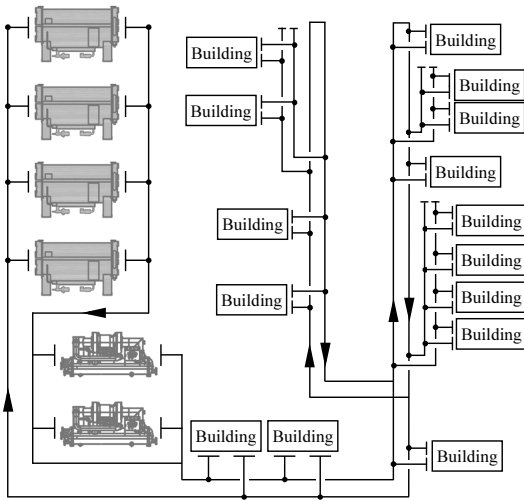
시스템의 내부 정보로 구성된다. 행동은 다중 냉동기의 ON/OFF 운영 조합으로 설계하였으며 행동에 따른 보상은 건물의 냉방부하 수요와 냉동기의 공급 부하 간의 차이로 설계하여 건물의 냉방부하와 근접한 의사결정을 하도록 하였다. 또한 냉동기 운영 제약으로 인해 발생하는 실현 불가능한 행동들을 방지하기 위한 훈련 알고리즘을 적용하였으며 이를 통해 학습 성능 및 안정성을 향상시켰다.

다중 냉동기 운영 모델은 실제 국내 A대학교 냉방시스템의 운영 데이터를 바탕으로 학습되었으며 학습에 사용되지 않은 데이터를 통해 운영 모델의 성능을 평가하였다. 실험을 통해 평가한 결과, 제안한 운영 모델이 최소가동시간을 준수함과 동시에 실제 캠퍼스의 냉동기 운영 로직보다 우수한 성능을 보임을 확인하였다. 이러한 결과는 본 연구 결과가 실제 냉방시스템에서의 운영 효율성을 향상시키는데 활용될 수 있음을 나타낸다.

본 연구의 구성은 다음과 같다. 2장에서는 고려하고자 하는 문제를 정의한다. 3장에서는 제안하는 냉동기 운영 모델의 프레임워크와 훈련 알고리즘에 대해 설명한다. 4장에서는 제안하는 냉동기 운영 모델과 기존 운영의 성능을 비교하였으며 마지막 5장에서는 결론 및 한계점을 제시한다.

II. 문제 정의

본 연구에서 고려하는 냉방시스템은 <Figure 1>과 같이 N 개의 냉동기를 운영하여 여러 건물에 냉방을 공급하는 중앙 냉방시스템이다. 해당 시스템은 건물의 냉방부하를 만족시키기 위해 냉동기 가동하여 냉방시스템을 순환하는 물의 온도를 낮추는 방식으로 냉방을 공급한다. 냉방 시스템 운영자는 각 냉동기의 ON/OFF 제어를 통해 건물에 공급하고자 하는 냉방부하 양을 결



<Figure 1> Structure of the central cooling system considered in this paper

정할 수 있다.

냉방시스템을 구성하는 냉동기는 서로 다른 정격용량을 가지며 Q_i 냉동기 i 의 정격용량을 나타내고 N 개의 냉동기로 구성된 냉방시스템은 총 2^N 개의 냉동기 운영 조합이 가능하다. 냉동기 운영 조합 의사결정은 Δt 간격으로 이루어지며 $O_{m,i}(t)$ 는 의사결정 t 시점에 결정되는 냉동기 운영 조합 m 의 냉동기 i 의 가동 상태를 의미하며 $O_{m,i}(t)$ 가 1일 경우 가동 상태를, 0일 경우 정지 상태를 의미한다. 냉동기 운영 의사결정에 따라 결정되는 t 시점의 냉방시스템의 공급부하는 $Q_{tot}(t)$ 로 정의되며 이는 식 (1)과 같이 계산된다. t 시점의 건물에서 발생하는 총 냉방부하 수요를 $B(t)$ 로 정의할 때, t 시점에 발생하는 공급과 수요의 차이 $D(t)$ 는 식 (2)와 같이 정의된다.

$$Q_{tot}(t) = \sum_{i=1}^N Q_i \cdot O_{m,i}(t) \quad (1)$$

$$D(t) = |Q_{tot}(t) - B(t)| \quad (2)$$

냉동기의 ON/OFF 변환이 빈번하게 발생할

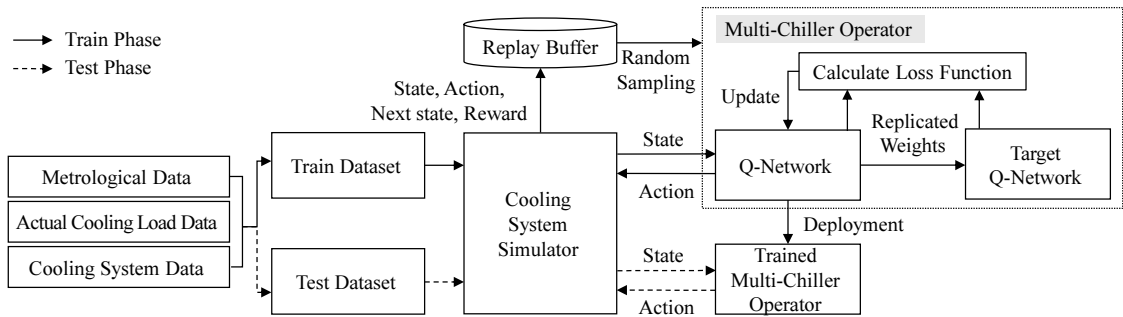
경우, 냉동기 내의 마모가 발생하며 수명이 단축되기 때문에 냉동기 운영에 있어 준수해야 할 최소가동시간이 존재한다. 본 연구에서 고려하는 문제의 경우, 모든 냉동기의 최소가동시간은 T_{min} 로 모두 동일하다. 이는 냉동기가 정지 상태에서 가동 상태로 전환될 경우, 반드시 T_{min} 동안 가동 상태를 유지해야 한다.

III. 제안 기법

3.1 DQN 기반의 다중 냉동기 운영 모델

다중 냉동기 운영 문제는 냉방시스템 내외부의 환경 요인들에 영향을 받으며 이러한 요인들과 냉동기 운영 간의 상호작용을 정확하게 모델링하기 어렵다. 이러한 문제를 해결하기 위해 본 연구는 전문 지식 없이 시행착오를 바탕으로 환경 변화에 따른 최적의 제어 동작을 추정할 수 있는 DQN 기반의 다중 냉동기 운영 모델을 제안한다. DQN은 환경과의 상호작용만으로 학습하는 모델 프리 방식으로 DQN 기반의 다중 냉동기 운영 모델은 다양한 환경에서의 제어 전략을 학습하고 이를 통해 환경 변화에 유연하게 대처 가능한 냉동기 운영 전략을 수립할 수 있다[12]. 운영 모델의 행동에 대한 상태 변화를 나타내기 위해 냉방시스템 시뮬레이터를 구축하였으며 이는 모델이 선정한 행동을 바탕으로 다음 시점의 냉방시스템의 환경 변화를 나타낸다.

DQN 기반의 다중 냉동기 운영 모델의 프레임워크는 <Figure 2>와 같다. 프레임워크는 학습 과정과 테스트 과정으로 구성된다. 학습 과정에서 냉방시스템 시뮬레이터는 실제 기후 정보, 냉방부하, 냉방시스템 데이터를 바탕으로 초기 상태를 구성한다. 냉방시스템 시뮬레이터는 의사결정시점의 냉방시스템의 내외부 환경 정보를 반영하는 상태를 Q-Network에 전달하고 Q-Network는 상태를 바탕으로 적절한 냉동기



〈Figure 2〉 Framework of proposed multi-chiller operation model

운영을 결정한다. Q-Network가 결정한 냉동기 운영 조합은 냉방시스템 시뮬레이터로 전달되며 결정된 냉동기 운영을 바탕으로 시뮬레이터의 다음 상태가 관측된다.

운영 모델에 의해 결정된 냉동기가 시뮬레이터 상에서 가동될 때, 시뮬레이터는 다음 상태를 관측함과 동시에 냉동기 운영에 따른 보상을 산출한다. 보상은 냉방부하의 과공급과 저공급을 절감시키도록 설계되었으며 건물의 냉방부하와 냉동기의 공급부하 간의 차이로 계산된다. 산출된 상태, 행동, 보상, 다음 상태는 리플레이 버퍼에 쌓이고 상호작용에 대한 경험은 모델 학습 시 활용된다. 리플레이 버퍼에 쌓인 경험들을 무작위 추출하여 학습할 경우, 데이터의 상관성을 줄여 학습의 안정성을 높일 수 있으며 경험 재활용을 통해 학습의 효율성을 높일 수 있다[14].

리플레이 버퍼에서 임의로 추출된 경험들은 Q-Network 학습을 위해 사용된다. 추출된 경험들을 바탕으로 Q-Network는 의사결정시점마다 현재의 Q값과 타겟 네트워크의 Q값의 차이를 나타내는 손실함수를 줄이는 방향으로 Q-Network의 가중치를 업데이트한다[17]. 타겟 네트워크는 Q-Network와 동일한 구조의 신경망으로 일정한 스텝마다 한 번씩 Q-Network의 가중치를 복제하는 방식으로 업데이트되며 학습의 안정성을 향상시킨다[9].

테스트 과정에서는 학습 과정에서 사용되지 않은 테스트 데이터들을 통해 학습된 냉동기 운영 모델의 성능을 평가한다. 테스트에는 학습된 두 네트워크 중 Q-Network가 사용된다. 학습된 냉동기 운영 모델은 시뮬레이터의 상태에 따른 냉동기 운영을 결정한다. 냉동기 운영 성능은 냉방시스템 가동 시간이 종료되는 시점까지의 과공급과 저공급을 기준으로 성능 평가가 이루어진다.

3.2 상태, 행동, 보상

다중 냉동기 운영을 위해 고려한 모델의 상태는 <Table 1>과 같다. 건물 외부의 기후 변화는 건물의 냉방부하 변화에 영향을 미치는 주요 요인으로 고려됨에 따라, 상태에 의사결정시점마다 관찰되는 외기 온도와 습도 정보를 반영하였다[10]. 시간에 따른 냉방부하 패턴의 학습을 위해 의사결정 시점 정보를 상태에 반영하였으며 냉방시스템의 내부 환경 변화를 반영할 수 있도록 시뮬레이터에 의해 도출되는 냉수의 환수온도와 공급온도를 상태로 활용하였다. 또한 건물의 냉방부하 수요에 대한 패턴 및 추세를 학습하도록 의사결정시점 이전의 한시간 동안의 측정되는 실제 건물의 냉방부하 수요를 상태에 반영하였다. 마지막으로, DQN 기반의 냉동기 운영 모델의 운영 기록과 냉동기 가동 상태 정보를 상태에 반영하여 운영 모델이 냉동기의 최소

〈Table 1〉 Feature of state considered at time t

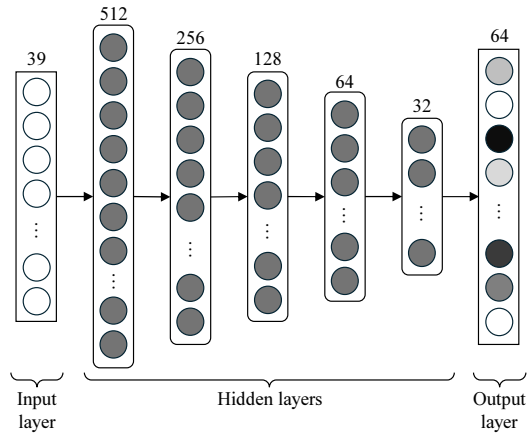
No.	Description	Categories	Size
1	Outdoor temperature	Observed value	1
2	Outdoor humidity	Observed value	1
3	Time	Observed value	1
4	Return temperature of chilled water	Obtained from Simulation	1
5	Supply temperature of chilled water	Obtained from Simulation	1
6	Actual total cooling load of buildings for last one hour	Observed value	4
7	Chillers operation status for last one hour	DQN action history	24
8	Chiller ON time duration	DQN action history	6

가동시간에 대한 제약을 학습하도록 설계하였다.

상태 정보를 바탕으로 다중 냉동기 운영 모델은 행동으로 의사결정 시점의 냉동기의 운영 조합을 결정한다. 결정된 냉동기 운영 조합은 냉방시스템 시뮬레이터 상에서 수행되며 행동에 따른 다음 상태와 보상이 업데이트 된다. 냉동기 운영에 따른 보상은 냉방 과공급과 저공급을 모두 절감시키기 위해 공급 부하와 건물의 냉방 부하 수요의 차이의 절댓값으로 정의하며 이는 식 (2)와 같이 계산된다.

3.3 다중 냉동기 운영 모델 구조

다중 냉동기 운영 모델의 Q 네트워크 구조는 <Figure 3>과 같이 1개의 입력 층, 5개의 은닉 층, 1개의 출력 층으로 구성된다. 네트워크의 입력 변수는 3.2장에서 정의된 상태가 입력되고 출력 층은 모든 냉동기 운영 조합에 대한 예측된 Q값을 산출하기에 입력층과 출력층은 각각 39, 64개의 노드로 구성된다. 은닉층은 각각 512, 256, 128, 64, 32개의 노드로 구성되며 복잡한 패턴 모델링과 기울기 손실 방지를 위해 활성화함수는 ReLU(Rectifier linear unit)를 적용하였다[18]. 은닉층과 출력층은 활성화 함수를 통해 이전 층의 출력 값들에 가중치를 곱하여 선형 결합한 값과 ReLU를 통해 계산된 비선형 값을 바탕으로 냉동기 운영 조합에 대한 Q값을 산출한다.



〈Figure 3〉 Q-network structure of the proposed model

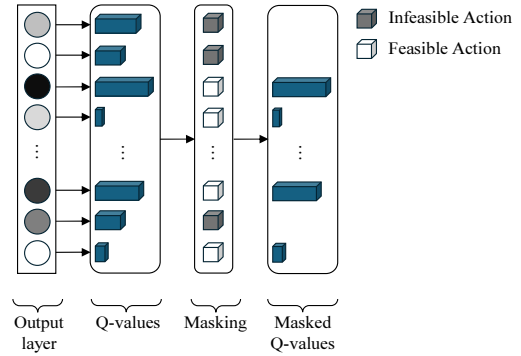
건물의 실제 냉방부하와 모델의 출력 값인 냉동기의 공급 부하 사이의 차이를 손실 값으로 정의하였으며 이를 최소화하는 방향으로 MSE (Mean Squared Error) 손실함수를 사용하여 가중치를 업데이트하였다. 모델 학습에 있어 네트워크 가중치 업데이트를 위한 옵티마이저로는 RMSProp를 사용하였다.

심층 강화학습 모델은 활성화 함수, 학습률, 네트워크 구조 및 크기 등과 같은 하이퍼파라미터 설정에 따라 성능의 차이가 존재한다[6]. 그러나 모든 하이퍼파라미터의 조합을 고려하여 최적의 성능을 갖는 하이퍼파라미터를 찾는 것은 해 공간이 넓고 매우 복잡한 문제이다. 따라

서 본 연구에서는 충분한 반복 실험을 통해 가장 우수한 성능을 보이는 하이퍼파라미터를 설정하였으며 이는 4.3장에서 다룬다.

3.4 학습 알고리즘

본 연구에서 제안하는 다중 냉동기 운영 모델은 냉동기의 최소화동시간 제약을 고려하기 위해 행동 마스킹 기반의 학습 알고리즘을 적용하였다. 냉동기 운영 시, 최소화동시간 제약으로 인해 실현 불가능한 운영 조합이 발생한다. 현실에서 발생할 수 없는 운영 조합들은 학습 과정에서 무의미함과 동시에 안정성을 저하시킨다. 이를 방지하기 위해 <Figure 4>와 같이 모든 냉동기 운영 조합에 대한 예측된 Q값이 존재할 때, 운영 의사



<Figure 4> Action masking in training

결정 시점의 상태에 따라 실행 불가능한 행동들의 Q값을 $-\infty$ 로 대체하여 마스킹을 진행하였다[7]. 실현 불가능한 행동 마스킹을 통해 운영 모델은 알고리즘 상에서 실제 냉방시스템에서 실현가능

Algorithm 1. Training Algorithm of Proposed Multi-chiller Operator

1. Initialize replay buffer D with size b
 2. Initialize network Q with random weights θ
 3. Initialize target network \hat{Q} with random weights θ^-
 4. **for** episode = 1, E **do**
 5. Initialize state
 6. **for** $t = 1, \dots, T$ **do**
 7. Observe state s_t and chiller operation status C_t
 8. Based on C_t generate and apply action mask
 9. Choose operation set \bar{a}_t according to ϵ -greedy strategy
 10. Execute operation set \bar{a}_t
 11. Get next chiller operation status C_{t+1} , next state s_{t+1} and reward r_t
 12. Store experiences $(s_t, \bar{a}_t, r_t, s_{t+1})$ in D
 13. **if** enough experiences in D **then**
 14. Sample random minibatch of M transitions $(s_j, \bar{a}_j, r_j, s_{j+1})$ randomly in D
 15. Calculate $y_j = r_j + \gamma \cdot \max_{a'} \hat{Q}(s_{j+1}, a'; \theta^-)$
 16. Calculate $L = \frac{1}{M} \sum_{j=0}^{M-1} (Q(s_j, \bar{a}_j; \theta) - y_j)^2$
 17. Update θ weights using $RMSProp$ by minimizing the loss L
 18. **end for**
 19. **end if**
 20. $s_{t+1} \leftarrow s_t, C_{t+1} \leftarrow C_t$
 21. **end for**
 22. Copy θ to θ^- every v steps
 23. **end for**
-

한 냉동기 운영 조합을 결정할 수 있으며 이를 통해 학습의 성능을 향상시킬 수 있다[21].

최소가동시간을 고려한 다중 냉동기 운영 모델의 학습 과정은 Algorithm 1과 같다. 학습을 위해 크기 b 인 리플레이 버퍼를 구축하고 Q네트워크와 타겟 Q네트워크가 각각 무작위 가중치 θ , θ^- 를 가지도록 초기화한다. 1일 냉동기 운영이 하나의 에피소드로 고려하였으며 에피소드가 E번 반복될 때까지 학습이 진행된다. 다중 냉동기 운영 모델은 행동 마스킹 결과와 의사결정시점의 상태를 바탕으로 냉동기 운영 조합을 결정한다.

운영 조합을 결정함에 있어 입실론 그리디 기법이 활용된다. 입실론 그리디 기법은 강화학습 모델에서 탐색과 활용의 균형을 조절하기 위한 기법으로 임의의 값이 입실론 값보다 작을 경우, 운영 모델은 가장 큰 Q값을 갖는 운영 조합을 선택하고 반대의 경우, 무작위 운영 조합을 탐험한다. 선택된 운영 조합을 바탕으로 다음 운영 시점의 상태와 보상이 결정되며 상태, 행동, 보상, 다음 상태가 리플레이 버퍼에 저장된다.

네트워크 가중치 θ 업데이트를 위해 리플레이 버퍼로부터 전이들이 미니배치 형태로 무작위 선정되며 하나의 미니배치는 M 개의 전이들로 구성된다. 손실 값은 line 16-17과 같이 타겟 네트워크에 의해 계산된 타겟 값 y_t 과 Q 네트워크에 의해 계산된 Q값의 평균 제곱 오차로 계산된다. 이때, y_t 은 미래 보상에 대한 감가율을 의미하며 *RMSProp*를 통해 손실값 L 을 최소화하는 방향으로 가중치를 학습한다. 타겟 Q 네트워크의 가중치 θ^- 는 에피소드가 끝나는 시점에 Q 네트워크 가중치 θ 로 업데이트된다.

3.5 냉방시스템 시뮬레이터

냉동기 운영에 따라 변화하는 냉방시스템의 온도 변화를 반영하기 위해 본 연구는 다중회귀 모델 기반의 냉방시스템 시뮬레이터를 구축하

<Table 2> Input variables of the prediction models for cooling system simulation

Variables	Description
x_1	Supply temperature of chiller water
x_2	Return temperature of chilled water
x_3	Outdoor temperature
x_4	Outdoor Humidity
x_5	Demand for cooling load in buildings
x_6	Supplying cooling load in chillers

였다. 냉방시스템 시뮬레이터는 냉동기 가동에 따른 냉수의 공급온도 예측 모델과 건물 순환을 마친 냉수의 순환 온도 예측 모델로 구성된다. 두 예측 모델의 입력 변수는 <Table 2>로 동일하다. t 시점의 냉동기 운영 정보, 건물의 냉방부하 수요, 냉방시스템의 환경 정보를 바탕으로 각 예측모델은 $t+1$ 시점의 공급온도와 환수온도를 예측한다.

시뮬레이터를 구성하는 공급온도 예측 모델과 환수온도 예측 모델의 회귀식은 각각 식 (3)과 식(4)와 같다. 이러한 냉방시스템 시뮬레이터는 다중 냉동기 운영 모델이 선택하는 냉동기 운영을 바탕으로 다음 냉방시스템의 상태를 구축할 때 활용된다. 냉방시스템 시뮬레이터는 실제 냉방시스템 운영 데이터를 사용하여 학습하였으며 학습에 사용하지 않은 데이터를 통해 시뮬레이터의 성능을 검증하였다. 시뮬레이터의 성능 평가 결과는 4.2장에서 설명한다.

$$y_{supply} = -0.8016x_1 - 0.0916x_2 + 0.0916x_3 + 0.0299x_4 - 0.0036x_5 - 0.0547x_6 + 0.0196 \quad (3)$$

$$y_{return} = -0.0385x_1 + 0.9469x_2 + 0.1128x_3 + 0.0461x_4 - 0.0993x_5 + 0.0089x_6 - 0.0213 \quad (4)$$

IV. 실험 및 평가

4.1 실험 환경

본 연구는 캠퍼스의 건물들에 냉방을 공급하기 위해 6대의 냉동기를 운영하는 A대학 중앙 냉방시스템 환경을 바탕으로 실험을 진행하였다. 냉방시스템은 평일 오전 9시부터 오후 9시까지 운영되며 냉방 운영자는 매 15분마다 냉동기 운영 조합을 결정하여 공급할 냉방부하를 결정할 수 있다. 냉방시스템은 서로 다른 정격 용량을 갖는 냉동기로 구성되어 있으며 실험에 사용된 냉동기 사양은 <Table 3>과 같고 각 냉동기가 90%의 효율로 가동될 때, <Table 4>는 가능한 냉동기 운영 조합의 예시를 나타낸다. 각 냉동기는 수명 유지를 위해 모든 냉동기는 가동 시 유지되

<Table 3> Specification of chillers considered in this paper

No.	Description	Number	Cooling capacity (RT)
1	Large turbo chiller	1, 2	475
2	Small turbo chiller	3, 4	180
3	Absolute chiller	5, 6	360

<Table 4> Example of chiller operation combinations

No.	State of chiller						Cooling capacity (RT)
	1	2	3	4	5	6	
1	OFF	OFF	OFF	OFF	OFF	OFF	0
2	ON	OFF	OFF	OFF	OFF	OFF	427.5
3	OFF	ON	OFF	OFF	OFF	OFF	427.5
4	OFF	OFF	ON	OFF	OFF	OFF	162.0
5	OFF	OFF	OFF	ON	OFF	OFF	162.0
	⋮						
61	ON	ON	ON	OFF	OFF	OFF	1719.0
62	ON	ON	ON	ON	OFF	ON	1539.0
63	ON	ON	ON	ON	ON	OFF	1539.0
64	ON	ON	ON	ON	ON	ON	1899.0

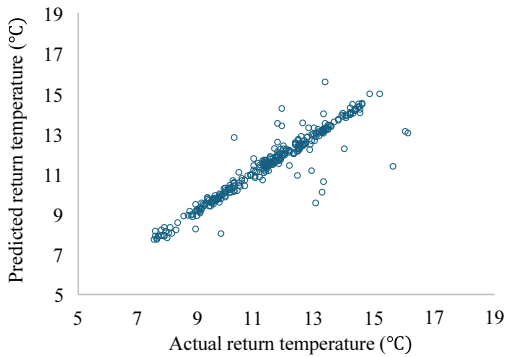
어야 하는 최소가동시간이 존재하며 고려하는 문제 상황의 최소 가동유지시간은 1시간이다.

제안한 다중 냉동기 운영 모델의 학습 및 평가를 위해 A대학 냉방시스템에 부착된 센서를 통해 계측된 냉방시스템 운영 데이터와 한국 기상청에서 제공하는 기상 데이터를 수집하였다. 데이터는 2021년 6월 1일부터 30일까지 15분 단위로 수집되었으며 냉방시스템이 가동되는 평일에 해당하는 22일 중 16일 데이터는 학습에 사용하고 나머지 6일은 테스트를 위해 사용하였다. 다양한 냉방 수요 패턴에 따른 성능을 평가하기 위해 테스트 데이터는 서로 다른 패턴을 보이는 일자 선정하여 구성하였다. 다중 냉동기 운영 모델의 성능은 테스트 기간에 해당하는 실제 냉방시스템 운영 기록을 바탕으로 평가하였으며 냉동기 운영을 통해 공급한 냉방부하와 건물의 냉방부하 사이의 차이를 바탕으로 운영 성능을 비교하였다.

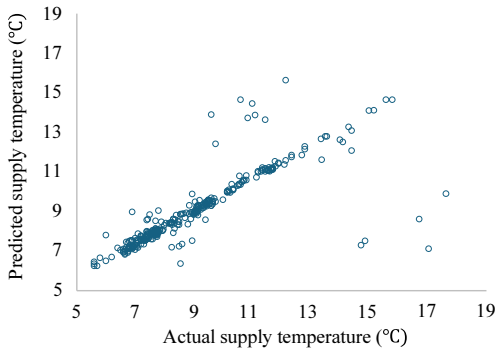
4.2 냉방시스템 시뮬레이터 검증

냉방시스템 시뮬레이터의 성능 검증은 학습에 사용되지 않은 데이터를 통해 수행되었으며 시뮬레이터를 구성하는 환수온도와 공급온도 예측 모델의 검증 데이터 셋에 대한 예측 성능은 회귀 모델에서 자주 사용되는 결정계수 (Coefficient of determination, R^2)를 통해 평가하였으며 이는 값이 1에 가까울수록 성능이 우수함을 의미한다. 환수온도 예측 모델은 $R^2 = 0.889$, 공급온도 예측 모델은 $R^2 = 0.666$ 으로 이러한 결과는 두 모델로 구성된 냉방시스템 시뮬레이터가 환경 변화에 따른 냉방시스템의 내부 환경 변동성 잘 반영하고 있음을 나타낸다.

검증 데이터셋에 대한 두 예측 모델의 결과는 각각 <Figure 5>와 <Figure 6>과 같다. 환수온도 예측 모델의 경우, 그래프 상의 점들이 경향성을 보이는 대각선에 밀집되어 있는 형태를 보이



<Figure 5> Prediction results obtained by the return temperature prediction model

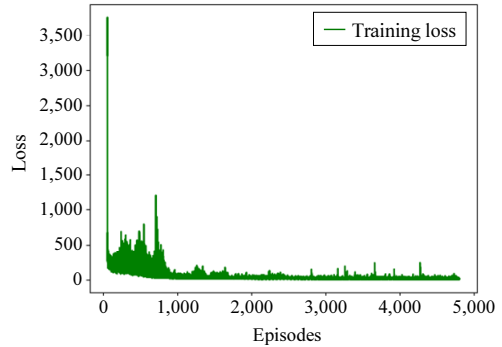


<Figure 6> Prediction results obtained by the supply temperature prediction model

며 예측 모델이 실제 값과 근사한 예측 값을 도출함을 나타낸다. 공급온도 예측 모델 역시 대부분의 점들이 경향성을 보이는 대각선에 밀집되어 있는 것으로 보아 공급온도 예측 모델 역시 실제 값과 유사한 예측 값을 도출해 내고 있음을 확인할 수 있으나 예측 값이 증가함에 따라, 잔차가 커지는 경향을 보인다. 이는 등분산성 가정을 만족하지 않는 것으로 보이며 추후 종속변수 로그변환 등의 기법 적용을 통해 예측 성능을 향상시킬 수 있을 것으로 기대된다.

4.3 학습 결과

<Figure 7>은 본 연구에서 제안한 모델의 학습 과정 동안 학습 데이터에 대해 계산된 손실



<Figure 7> Training loss graph

<Table 5> Hyper-parameters for training

Hyper-parameters	Values
Batch size	128
Learning rate	0.001
Optimizer	RMSprop
Epsilon decay weight	0.999
Episode	5000
Replay buffer size	10,000
Target network update	50

값의 변화를 나타낸다. 학습 오류(Training Loss)는 학습 초기의 급격한 감소 이후 에피소드가 1,000번 반복되기 이전까지 크게 변동하는 모습을 보이지만 충분한 에피소드 반복 이후에는 큰 변동 없이 안정적으로 학습되는 것을 확인할 수 있다. 이러한 변화를 통해 제안한 모델이 초기 다양한 상태에서의 냉동기 운영 조합에 따른 상태 변화를 경험하고 경험을 통해 냉방시스템 내 외부의 환경 변화에 따른 적절한 냉동기 운영 조합을 선택하도록 학습되는 것을 확인할 수 있다.

<Table 5>는 냉동기 운영 모델 학습을 위해 사용된 하이퍼파라미터를 나타낸다. Du, Yan, et al.[14]에서 제시한 설정 값을 바탕으로 초기 모델을 구축하였으며 충분한 반복 실험을 통해 본 연구에서 고려하는 문제에서 가장 우수한 성능을 보이도록 하이퍼파라미터 값을 설정하였다. 모델 훈련은 32GB RAM, NVIDIA GeForce

RTX 3070, 11th Gen Intel(R) Core™ i7-11700F @ 2.50GHz가 장착된 PC에서 Python 언어를 통해 진행하였다.

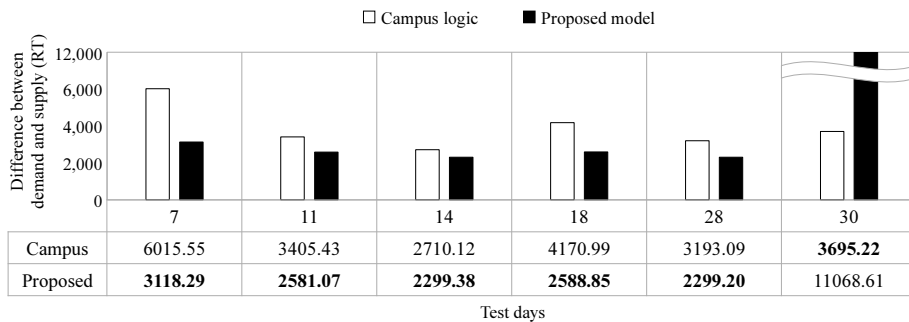
4.4 실험 결과

심층 강화학습 기반의 다중 냉동기 운영 모델의 성능을 검증하기 위해 제안한 냉동기 운영과 실제 A대학에서 사용하고 있는 냉동기 운영 로직을 비교하였다. 캠퍼스의 실제 운영 로직의 경우, 운영자의 노하우를 바탕으로 냉동기 운영 조합이 결정되며 의사결정은 2시간 간격으로 이루어진다.

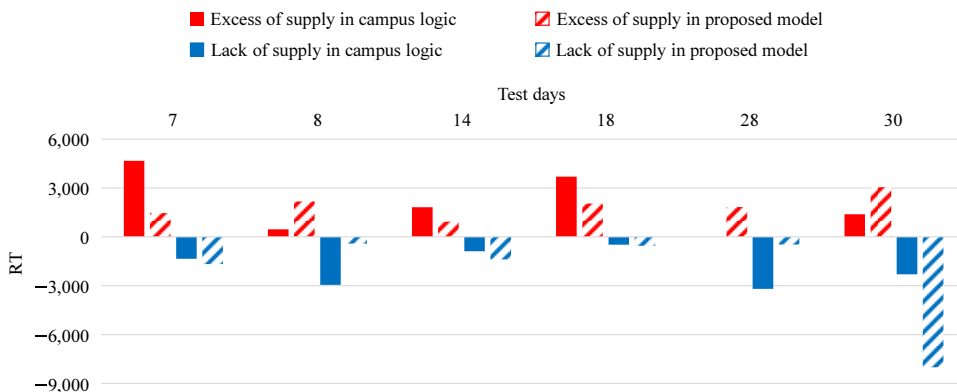
<Figure 8>은 학습에 사용되지 않은 날짜를 대상으로 수행된 건물의 냉방 수요와 공급 부하의 차이 측정에서 두 로직의 성능 비교 결과를 나타낸다. 30일을 제외한 모든 테스트 날짜에서

제안한 모델이 기존의 캠퍼스 운영 로직보다 우수한 성능을 보이며 약 30% 개선율을 확인할 수 있다. 그러나 30일의 경우, 다른 날짜들과 달리 냉방 운영 초기에 높은 냉방부하 수요가 발생하는 예외적인 냉방부하 패턴으로 인해 제안 모델이 낮은 성능 보인다. 이러한 결과는 제안된 냉동기 운영 모델이 학습 데이터셋에 포함된 패턴에서는 우수한 성능을 보이지만 학습 데이터와 다른 패턴의 냉방 수요 패턴에서는 성능이 저하됨에 따라 추가적인 냉방 수요 패턴 수집을 통해 제안한 모델의 성능을 향상시킬 수 있음을 나타낸다.

다중 냉동기 운영 결과를 과공급과 저공급 관점에서 분석한 결과는 <Figure 9>와 같다. 기존 캠퍼스의 냉동기 운영 로직은 대부분의 날짜에



<Figure 8> Results of chiller operation models in terms of difference between supply cooling load and demand of buildings



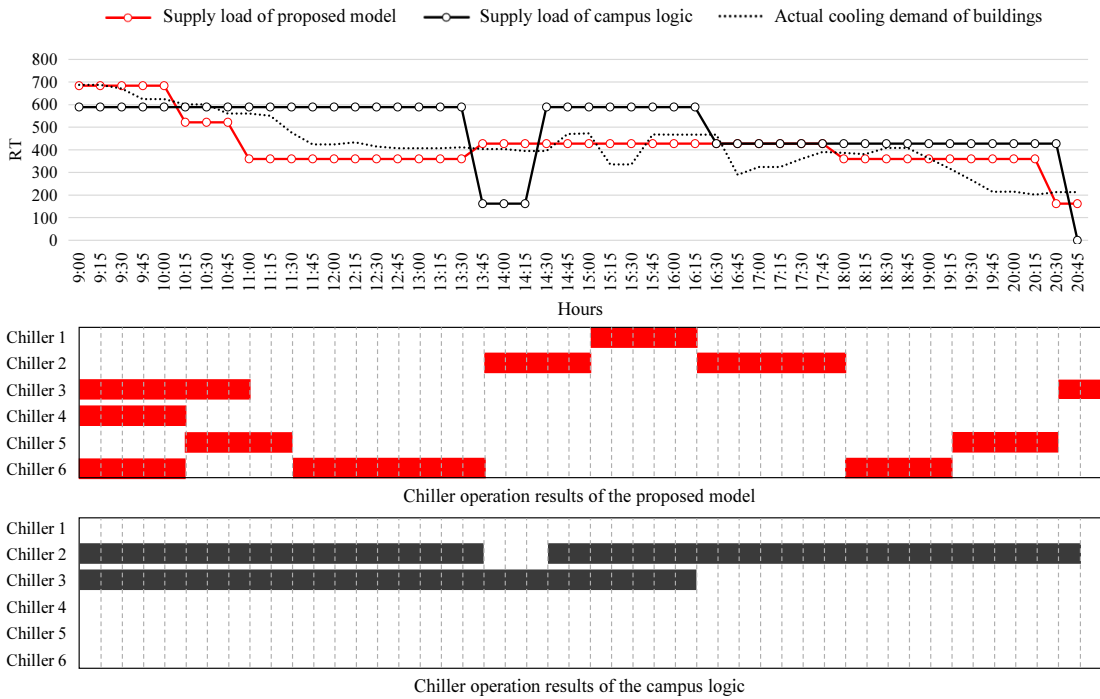
<Figure 9> Results of chiller operation models in terms of excess and lack of supply

서 과공급이나 저공급에 치우친 운영 결과를 보이는 반면 제안한 모델은 실제 건물에서 발생하는 냉방부하와 근사한 냉동기 운영을 통해 보다 균등한 결과를 보인다. 이러한 결과는 제안한 모델이 기존 냉동기 운영 로직의 과공급과 저공급으로 인해 발생할 수 있는 에너지 낭비와 실내 쾌적성 저하 문제를 완화할 수 있음을 시사한다. 30일의 경우, 다른 날짜들에 비해 높은 초기 냉방부하 수요가 발생함으로 인해 높은 저공급이 발생함을 확인할 수 있다.

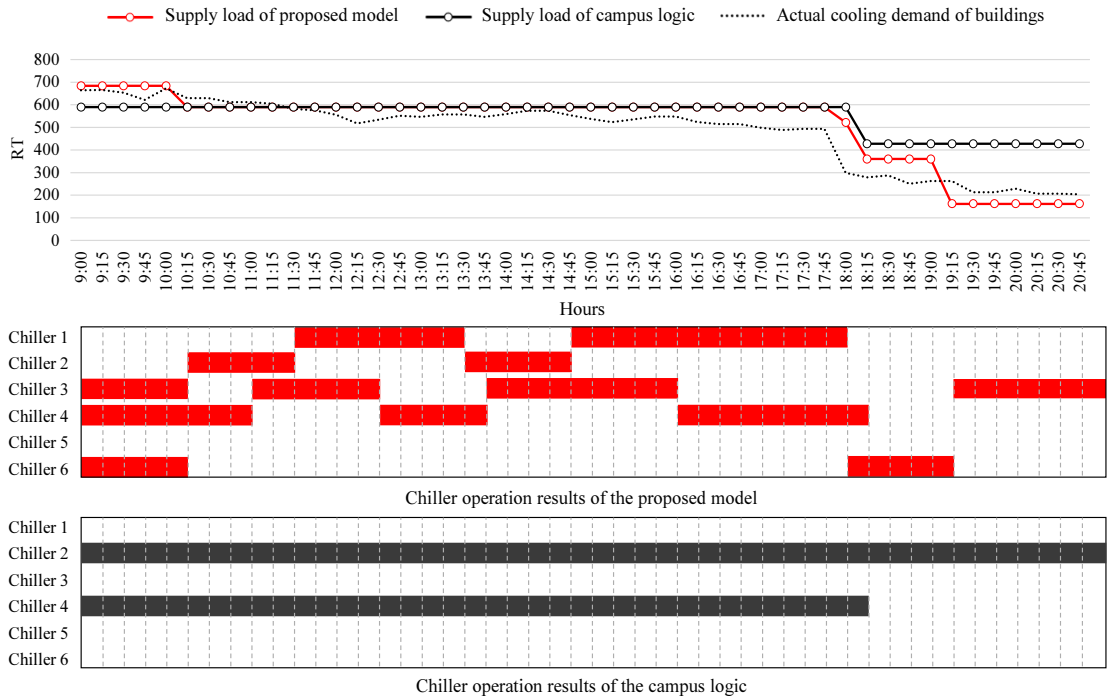
건물들의 냉방 수요 패턴에 따른 운영 성능을 비교하기 위해 제안한 모델과 캠퍼스의 실제 운영 결과를 서로 다른 패턴을 보이는 테스트 날짜인 7, 18, 30일에 대해 공급 부하 그래프와 간트 차트를 활용하여 나타내었다. <Figure 10>은 건물들의 총 냉방부하 값이 시간에 따라 꾸준히 감소하는 패턴에 대한 운영 결과를 나타낸다. 건물의 실제 부하와 공급 부하를 나타낸 그래프

를 통해 제안한 모델이 해당 패턴에서 실제 캠퍼스 운영 로직보다 건물의 냉방부하 수요를 잘 따라 가는 모습을 확인할 수 있으며 간트 차트를 통해 연속가동시간 제약 역시 잘 준수됨을 확인할 수 있다.

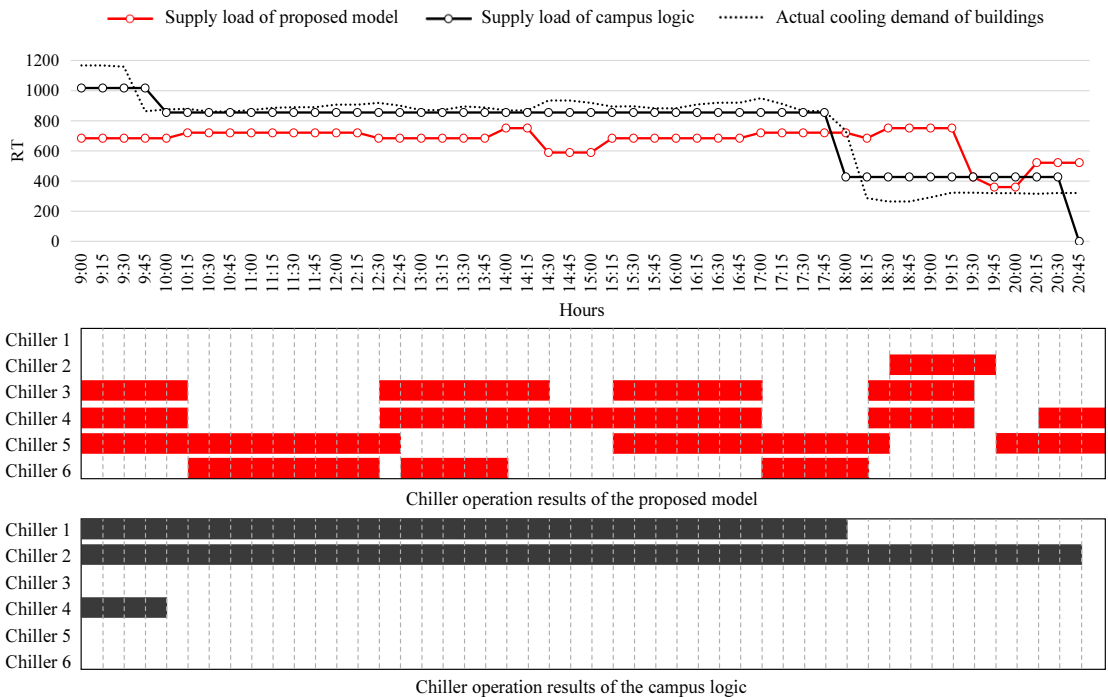
<Figure 11>와 <Figure 12>는 각각 18일과 30일에 대한 테스트 결과를 나타낸다. 18일은 건물들의 냉방 수요가 일정한 값을 유지하다 특정 시점 이후에 급격하게 감소하는 패턴을 보인다. 제안한 모델은 건물의 실제 부하를 잘 따라가는 것을 확인할 수 있으며, 특히, 냉방부하 수요가 급격하게 감소하는 시점이 잘 학습되었음을 확인할 수 있다. 30일은 다른 날짜와 달리 건물의 냉방부하 수요가 처음부터 높은 값을 보이는 상이한 패턴을 보인다. 해당 날짜에서 제안된 모델이 다른 날짜와 유사하게 700RT 이하의 냉동기 운영을 결정하며 저공급을 발생시킨다. 이러한 결과는 학습 시 제공되지 않은 패턴에 대해



<Figure 10> Operation results and Gantt chart for test day 7



〈Figure 11〉 Operation results and Gantt chart for test day 18



〈Figure 12〉 Operation results and Gantt chart for test day 30

서는 제안한 모델의 운영 성능이 떨어질 수 있으며 이에 따라 다양한 냉방 수요 패턴에 대한 학습이 요구됨을 나타낸다.

V. 결론

5.1 논문의 결론

본 연구에서는 중앙 냉방시스템 환경에서의 냉동기의 최소가동시간을 고려하기 위한 심층 강화학습 기반의 다중 냉동기 운영 모델을 제안하였다. 건물의 냉방부하 수요와 냉방시스템의 공급 부하 간의 차이를 최소화하는 냉동기 운영 조합을 학습하기 위한 상태와 보상을 설계하였으며 실제 냉방시스템을 모사한 시뮬레이터를 통해 결정된 냉동기 운영 조합에 따른 상태 변화를 관측하였다. 또한 행동 마스킹을 적용한 훈련 알고리즘을 적용함으로써 냉동기 운영 모델이 실현 불가능한 냉동기 운영 조합이 선택하는 것을 방지하였다. 제안한 심층 강화학습 기반의 다중 냉동기 운영 모델의 성능을 평가하기 위해 A대학 실제 냉방시스템의 냉동기 운영 로직과 성능 비교 실험을 진행하였다.

테스트 날짜를 대상으로 실험한 결과, 30일을 제외한 모든 날짜에서 기존 캠퍼스 운영 로직보다 건물의 냉방부하와 근접한 운영을 통해 과공급과 저공급을 절감시킴을 확인하였다. 30일의 경우, 기존 날짜들에 비해 냉방 수요가 급격하게 상승한 패턴으로 6월에 한정된 학습 데이터 내에서 비중이 매우 적다. 따라서 제안 모델이 해당 패턴에 대한 충분한 학습을 거치지 못한 것으로 예상되며 이는 추후 냉방이 가동되는 7, 8월에 해당하는 다양한 냉방 수요 패턴의 데이터 수집을 통해 성능을 개선할 수 있으리라 기대된다. 또한, 모든 날짜에서 제안 모델이 냉동기의 연속가동시간을 준수하며 실행 가능한 운영 조합을 결정하는 것을 확인하였으며 이러한

결과는 학습된 다중 냉동기 운영 모델이 실제 연속가동시간이 존재하는 냉방시스템에서 활용될 수 있을 것이라 기대된다. 더 나아가 제안한 모델이 실제 운영 환경에서 활용될 시, 냉방시스템 에너지 소비 최적화, 운영 비용 절감 등 다방면의 의사결정에 기여할 수 있을 것이라 기대된다.

5.2 본 논문의 한계점 및 향후 연구방향

다만 4.4장에서 언급한 바와 학습 데이터에 포함되지 않거나 비중이 작은 냉방부하 패턴에 대해서는 성능 저하가 발생한다는 한계점이 존재한다. 추가적인 데이터 수집이나 다양한 패턴의 데이터 증강 연구를 통해 학습 데이터 내 패턴의 다양성을 높임으로써 모델의 강건성을 높이기 위한 노력이 필요하다. 또한 냉동기 운영에 따른 냉방시스템의 환경 변화를 반영하기 위한 시뮬레이터 성능 고도화 연구가 요구된다. 실제 운영에서 발생할 수 있는 다양한 상황에 대처하기 위해 시뮬레이터 고도화가 필요하며 이를 통해 운영 모델의 안정성을 향상시킬 수 있을 것이다.

참 고 문 헌

- [1] Beghi, A., Cecchinato, L., Cosi, G., and Rampazzo, M. (2012). "A PSO-based algorithm for optimal multiple chiller systems operation" *Applied Thermal Engineering*, 32, pp. 31-40.
- [2] Chang, Y. C. (2006). "An outstanding method for saving energy-optimal chiller operation" *IEEE transactions on energy conversion*, 21(2), pp. 527-532.
- [3] dos Santos Coelho, L., Klein, C. E., Sabat, S. L., and Mariani, V. C. (2014). "Optimal chiller

- loading for energy conservation using a new differential cuckoo search approach” *Energy*, 75, pp. 237-243.
- [4] Du, Y., Li, F., Kurte, K., Munk, J., and Zandi, H. (2022) “Demonstration of intelligent HVAC load management with deep reinforcement learning: real-world experience of machine learning in demand control” *IEEE Power and Energy Magazine*, 20(3), pp. 42-53.
- [5] He, K., Fu, Q., Lu, Y., Wang, Y., Luo, J., Wu, H., and Chen, J. (2023). “Predictive control optimization of chiller plants based on deep reinforcement learning” *Journal of Building Engineering*, 76, pp. 107158.
- [6] Henderson, P., Islam, R., Bachman, P., Pineau, J., Precup, D., and Meger, D. (2018). “Deep reinforcement learning that matters” In *Proceedings of the AAAI conference on artificial intelligence*, Louisiana, USA, 32(1).
- [7] Huang, S., and Ontañón, S. (2020). “A closer look at invalid action masking in policy gradient algorithms” *arXiv preprint arXiv:2006.14171*.
- [8] Hyun, S. H., Park, C. S., and Augenbroe, G. L. M. (2008). “Analysis of uncertainty in natural ventilation predictions of high-rise apartment buildings” *Building Services Engineering Research and Technology*, 29(4), pp. 311-326.
- [9] Jia, W., Li, J., and Zhao, Y. (2021). “DQN Algorithm Based on Target Value Network Parameter Dynamic Update” *IEEE 4th International Conference on Computer and Communication Engineering Technology (CCET)*, Beijing, China, pp. 285-289.
- [10] Kwok, S. S., Yuen, R. K., and Lee, E. W. (2011). “An intelligent approach to assessing the effect of building occupancy on building cooling load prediction” *Building and Environment*, 46(8), pp. 1681-1690.
- [11] Kwon, Y. D., Choo, J., Kim, B., Yoon, I., Gwon, Y., and Min, S. (2020). “Pomo: Policy optimization with multiple optima for reinforcement learning” *Advances in Neural Information Processing Systems (NeurIPS)*, Vancouver, Canada, 33, pp. 21188-21198.
- [12] Lewis, F. L., and Vrabie, D. (2009). “Reinforcement learning and adaptive dynamic programming for feedback control” *IEEE circuits and systems magazine*, 9(3), pp. 32-50.
- [13] Liao, Y., and Huang, G. (2019). “A hybrid predictive sequencing control for multi-chiller plant with considerations of indoor environment control, energy conservation and economical operation cost” *Sustainable Cities and Society*, 49, pp. 101616.
- [14] Liu, R., and Zou, J. (2018). “The effects of memory replay in reinforcement learning” In *2018 56th annual allerton conference on communication, control, and computing (Allerton)*, IEEE, Monticello, USA, pp. 478-485.
- [15] Liu, Z., Tan, H., Luo, D., Yu, G., Li, J., and Li, Z. (2017). “Optimal chiller sequencing control in an office building considering the variation of chiller maximum cooling capacity” *Energy and Buildings*, 140, pp. 430-442.
- [16] Mavrovouniotis, M., Li, C., and Yang, S. (2017). “A survey of swarm intelligence for dynamic optimization: Algorithms and applications” *Swarm and Evolutionary Computation*, 33, pp. 1-17.
- [17] Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., and Riedmiller, M. (2013). “Playing atari with deep reinforcement learning” *arXiv preprint arXiv:1312.5602*.
- [18] Nair, V., and Hinton, G. E. (2010). “Rectified

linear units improve restricted boltzmann machines” In Proceedings of the 27th international conference on machine learning (ICML-10), Haifa, Israel, pp. 807-814.

- [19] Sala-Cardoso, E., Delgado-Prieto, M., Kampouropoulos, K., and Romeral, L. (2020). “Predictive chiller operation: A data-driven loading and scheduling approach” *Energy and Buildings*, 208, pp. 109639.
- [20] Wong, W., Dutta, P., Voicu, O., Chervonyi, Y., Paduraru, C., and Luo, J. (2022). “Optimizing industrial hvac systems with hierarchical reinforcement learning” *arXiv preprint arXiv:2209.08112*.
- [21] Wu, Z., Li, Y., Zhan, W., Liu, C., Liu, Y. H., and Tonizuka, M. (2024). “Efficient Reinforcement Learning of Task Planners for Robotic Palletization through Iterative Action Masking Learning” *arXiv preprint arXiv:2404.04772*.
- [22] Ye, Y., Sharma, R., and Guo, F. (2016). “A multi-layer optimal chiller operation management framework” *IEEE Industry Applications Society Annual Meeting (IAS)*, Portland, USA, pp. 1-6.
- [23] Zhuang, C., Wang, S., and Shan, K. (2020). “A risk-based robust optimal chiller sequencing control strategy for energy-efficient operation considering measurement uncertainties” *Applied Energy*, 280, pp. 115983.

저자 소개



김 종 은(Jongeun Kim)

- 2022년 : 인천대학교 산업경영공학과(학사)
- 2022년~현재 : 인천대학교 산업경영공학과(석사)
- <관심분야> : 딥러닝, 빅데이터 분석, 강화학습, 소셜네트워크 분석



김 관 호(Khanho Kim)

- 2016년 : 동국대학교 정보시스템전공(학사)
- 2012년 : 서울대학교 산업공학과(석사)
- 2013년 : 경희대학교(연구박사)
- 2023년: 인천대학교 산업경

영공학과 교수

- 2024년~현재: 동국대학교 산업시스템공학과 교수
- <관심분야> 딥러닝, 빅데이터분석, 스마트팩토리, 강화학습



김 재 곤(Jae-Gon Kim)

- 1994년 : 한국과학기술원 산업 및 시스템공학과(학사)
- 1996년 : 한국과학기술원 산업 및 시스템공학과(석사)
- 2001년 : 한국과학기술원 산업 및 시스템공학과(박사)

- 2005년~현재: 인천대학교 산업경영공학과 교수
- <관심분야> 산업인공지능, 데이터기반 운영최적화