Sensors and Materials, Vol. 36, No. 9 (2024) 3917–3931
MYU Tokyo

S & M 3775

Multi-residential Heating, Ventilation and Air Conditioning Control Based on Deep Reinforcement Learning

Seunghoon Lee*

Department of Industrial Management Engineering, Dong-A University,
37, Nakdong-Daero 550 beon-gil, Saha-gu, Busan 49315, Republic of Korea

(Received June 11, 2024; accepted September 5, 2024)

Keywords: HVAC system control, deep reinforcement learning, control optimization, energy efficiency, sensor application

Improving heating, ventilating, and air conditioning (HVAC) efficiency is crucial for energy savings and carbon emission reduction. In this study, we employed deep reinforcement learning (DRL) to optimize HVAC system control in commercial buildings. Traditional control methods, such as rule-based and model predictive control, often fall short in dynamic and complex environments. In contrast, DRL combines reinforcement learning with deep neural networks to provide a more adaptive and efficient approach. Focusing on a multi-floor commercial building, we used a binary on/off control strategy to streamline decision-making and enhance scalability. The HVAC control problem is modeled as a finite Markov process, with a deep Q-network optimizing operations based on parameters such as indoor/outdoor temperatures, cloud coverage, and occupancy levels. A comparative analysis using simulations and real-world data collected by sensors from a commercial building in South Korea showed that the DRL-based method significantly reduced the HVAC operation frequency and on/off cycles, achieving superior energy savings while maintaining comfortable temperature levels. These results highlight the potential of DRL for effective HVAC management by balancing energy efficiency with occupant comfort.

1. Introduction

Building energy consumption constitutes a substantial portion of global energy use and has a rapid upward trajectory. It currently accounts for 40% of the global primary energy consumption, contributing significantly to 30% of CO_2 emissions.[1] Among the various building services, heating and air conditioning systems, which are essential for maintaining optimal indoor temperatures for occupant comfort, are the most energy intensive. In recent years, abnormal weather conditions have further increased energy usage, which is largely driven by the demand for heating and cooling. Enhancing the efficiency of energy control systems is crucial to address this challenge. Such improvements are essential not only for realizing significant energy savings in building operations, but also for reducing carbon emissions.

*Corresponding author: e-mail: seungh@dau.ac.kr
https://doi.org/10.18494/SAM5183

ISSN 0914-4935
https://myukk.org/

Technological advancements such as cloud computing and artificial intelligence (AI) facilitate real-time data monitoring and communication with devices, enabling instantaneous decision-making. In the building sector, these innovations seamlessly integrate intelligence and control systems, culminating in the development of a unified and efficient system for building operations: the smart building management system.[2] This system optimizes building energy use by collecting internal and external weather information through Internet-of-Things devices, thereby allowing the real-time control of heating and air conditioning systems. Furthermore, occupant movements within buildings are tracked via their cell phones connected to beacon receivers, enabling adaptive adjustments to indoor temperatures to address congestion and enhance comfort. To enhance the effectiveness of building management systems further, incorporating parameters that are associated with occupants, including behavior, preferences, and interactions with the building, into the control algorithm is crucial. The predominant focus of these systems is on balancing two crucial factors: optimizing occupant comfort and achieving energy-saving goals.[3] This balance is particularly important in commercial buildings, where energy use directly affects costs and occupant comfort influences sales and space utilization.

Numerous studies have delved into heating, ventilating, and air conditioning (HVAC) systems, underscoring their pivotal role in overall building energy consumption. Strategies for enhancing energy efficiency include upgrading outdated systems and integrating natural energy sources. Effective control methods are essential to achieve energy savings while ensuring occupant comfort. Rule-based and model predictive control (MPC) methodologies are often recommended for efficient HVAC system control. Rule-based control is simple and effective in static situations. However, its performance may deteriorate in dynamic scenarios. Conversely, MPC, which is designed using optimization techniques, ensures optimal operation but can be time-consuming and sometimes infeasible in complex situations. Addressing the uncertainty in dynamic scenarios poses challenges for both rule-based control and MPC.

In recent years, AI techniques such as deep reinforcement learning (DRL) have been employed to enhance energy efficiency while simultaneously ensuring occupant comfort. DRL combines RL and deep learning using a neural network (NN). In the RL framework, an agent interacts with the environment, observes the state, and selects an action from a predefined set. The environment provides a reward based on the selected action. In sequential decision-making, an NN guides the action selection. Through iterative processes, the objective of RL is to train the agent to maximize its rewards in a given environment.

The complexity of HVAC systems is notable, encompassing various components such as heating equipment, controller systems, and refrigerants that work in tandem to regulate the indoor environment of a building. These systems are typically found in large or new commercial buildings with sophisticated centralized control mechanisms and are less prevalent in smaller commercial establishments. Tenants in smaller or older buildings often use individual HVAC systems such as heat pumps and air conditioners, which are simpler than those in large buildings. The operation of individual HVAC units, which are controlled by occupants, introduces fluctuations in the energy usage according to individual preferences, making energy efficiency susceptible to occupant behavior.

In this paper, an HVAC system control algorithm based on DRL is presented to improve energy efficiency and maintain occupant comfort. The algorithm manages the status of the HVAC system through on and off actions to regulate the indoor temperature effectively while conserving energy. In addition, a separate deep NN is developed to predict the indoor temperature based on the status of the HVAC system. The performance of the algorithm was compared with that of a rule-based control approach, which was evaluated in a case study with the real data collected by sensors.

The remainder of this paper is organized as follows. In Sect. 2, we present a literature review of the control methodologies for HVAC systems in buildings and explore DRL. The details of the building model and the DRL algorithm using the Markov decision process are described in Sect. 3. The details of the experiments with the DRL algorithm in comparison with rule-based methods are presented in Sect. 4. Finally, in Sect. 5, we conclude the paper with a summary of the findings.

2. Literature Review

As the focus on climate and energy conservation intensifies, numerous studies have focused on improving energy efficiency in the building sector, paying special attention to HVAC systems, which constitute a significant portion of the energy consumption in buildings. Rule-based methodologies that incorporate information relating to building occupants have been introduced to control HVAC systems for energy savings. Agarwal *et al.* devised a control system that orchestrates the activation and deactivation of HVAC systems based on the detection of the presence and absence of occupants using passive infrared (PIR) and door sensors.[4] Another study explored the manipulation of cooling and heating setpoints through the connection of smartphones to Wi-Fi infrastructure, contingent on occupant detection within rooms.[5] Padmanabh *et al.* introduced control logic with the aim of increasing energy efficiency in conference spaces.[6] The occupancy status was meticulously determined by the controller, which set a threshold for measurements acquired from light and sound sensors. Consequently, the HVAC system and lighting were deactivated during unoccupied periods.[6]

In another study, an occupancy prediction model using a particle filter was proposed to determine precisely the current occupancy status of a building. Data collected from PIR sensors and cameras were used for occupancy estimation, and energy was saved through adjustments to ventilation rates and room-temperature target points based on occupancy measurements and predictive analytics.[7] Gao and Keshav proposed a model for predicting future room temperatures based on the current power of the HVAC system.[8] They identified an optimal control strategy and determined the most efficient time to activate or deactivate the HVAC system based on the model. For example, the HVAC system was activated 10 min before the arrival of the occupant and deactivated earlier than the expected departure time, thereby showcasing an energy-saving approach.[8] Li *et al.* introduced two control strategies to optimize HVAC systems for energy savings.[9] They addressed the challenges associated with the inclination to set a lower temperature setpoint and the tendency to forget to turn off the HVAC system. These issues were resolved by implementing anomaly detection and the automatic on/off control of the HVAC system.[9]

MPC has also been employed in smart building management to manage HVAC systems. MPC involves making instantaneous decisions to control commands at every moment within a planning horizon by solving an optimization problem.[10] Dong and Lam introduced a nonlinear MPC approach for HVAC system control by incorporating the anticipation of occupant behavior patterns and weather conditions.[11] Indoor environmental parameters, power consumption, and ambient conditions were monitored using sensors, and occupant behavior patterns were predicted using Markov models. By leveraging this information, temperature adjustments were made in unoccupied zones, thereby contributing to energy efficiency.[11] Široký *et al.* similarly employed MPC integrated with weather predictions for the HVAC system with the aim of minimizing energy consumption.[12] This approach involved reducing the room temperature during nights and weekends.[12] Another study used a model predictive controller with the prediction of zone loads and weather conditions to optimize the operational efficiency of HVAC systems.[13]

In recent years, active discourse has arisen on research employing AI techniques. Esrafilian-Najafabadi and Haghighat introduced an HVAC control system using a deep learning algorithm to predict the preheating time and occupancy patterns.[14] Their study facilitated energy reduction by regulating the setback and setpoint temperatures.[14] Other studies explored HVAC system control with a focus on occupancy patterns using machine learning techniques.[15,16] RL, which is another AI technique, has also been applied to controlling HVAC systems. In a study by Wei *et al.*, DRL was employed to address variable airflow volume control within an HVAC system.[17] Similarly, Brandi *et al.* applied DRL to optimize energy savings in an HVAC system, where the agent selected one of the suggested supply water temperature setpoints.[18] Another study utilized the DRL framework, considering both the adjustment of the thermostat and occupant behavior regarding clothing choices.[19] Wang *et al.* employed a model-free actor–critic DRL algorithm to optimize the thermal comfort and energy consumption of HVAC systems by controlling the setpoints.[20] Ahn and Park controlled the setpoints of HVAC systems using a deep Q-network (DQN), which is another DRL algorithm, to strike a balance between different HVAC systems.[21] DRL methodologies have exhibited better performance in saving energy and maintaining occupant comfort than other methods.[21]

Although existing research has predominantly focused on using AI techniques, especially DRL, to control HVAC system setpoints, notable challenges impact the practical implementation of such approaches. The issue of exponential complexity becomes evident when considering a higher number of temperature setpoints, introducing computational challenges in exploring and optimizing the vast action space. This challenge is particularly relevant in real-world scenarios in which multiple room zones need to be controlled. The computational demands associated with training DRL methods, coupled with the exponential growth in complexity when managing multiple zones, pose significant hurdles to scalability and real-world feasibility.[22]

The distinctiveness of this study lies in its contribution to overcoming the well-known scalability barrier in DRL applications for HVAC systems. Unlike previous methods that struggle with the exponential complexity of multi-zone control, our binary representation effectively reduces the computational burden, enabling the DRL method to be more easily implemented in real-world settings. Moreover, in this study, we demonstrated that even with a

simplified action space, the DRL-based approach retains the ability to optimize energy efficiency and maintain thermal comfort across multiple zones. This balance between simplicity and effectiveness highlights the potential for the widespread application of our method in diverse building environments, paving the way for more scalable and practical AI-driven HVAC control systems.

3. HVAC System Control Problem

In this study, a commercial building with multiple stories was selected as the application domain for implementing DRL to control HVAC systems. The distinctive feature of this commercial building lies in the paramount importance of occupant comfort, given its direct impact on sales. Nevertheless, energy conservation remains a crucial objective. The HVAC systems regulate the indoor temperature on each floor, offering different modes such as "Cooling," "Heating," and "Auto." In the "Auto" mode, the HVAC system can autonomously switch between cooling and heating on the basis of the indoor temperature and specified setpoint. For generality, we focused on scenarios in which all floors required cooling. The primary objective of the HVAC system control was to maintain the indoor temperature within the user comfort range while minimizing energy costs.

3.1 DRL-based HVAC system control in commercial building

DRL has recently emerged as a focal point for researchers and practitioners, finding application across various industrial domains including the building sector. Within the realm of DRL, various algorithms have been developed, with the DQN being notable and still gaining traction. In contrast to the traditional Q-learning introduced by Watkins and Dayan, the DQN integrates an NN into its framework.[23] The DQN, which was proposed by Mnih *et al.*, distinguishes itself by approximating the Q-value through the NN and updating the network weights using a combination of replay memory and a target network.[24]

Unlike conventional Q-learning, in which Q-values are computed and updated in a Q-table based on state–action pairs, the DQN leverages an NN to generate an approximate Q-value. The replay memory stores the samples and a random subset of these samples is drawn for learning. This approach effectively addresses the coupling characteristics between learning data, ensures stable learning, and mitigates overfitting. In addition, the inclusion of a target network plays a pivotal role in stabilizing the learning process by calculating the target value with a fixed parameter for specific steps. This prevents the target value from undergoing frequent changes, ultimately contributing to the stability of the learning process.

Numerous studies have employed DRL to regulate HVAC systems to curtail energy consumption while ensuring occupant comfort.[25–27] Given the intrinsic characteristics of buildings, optimizing these two objectives is pivotal for HVAC system control. Efforts have traditionally focused on manipulating the temperature setpoints of HVAC systems using DRL algorithms. This involves designing a set of candidate actions, each representing a specific temperature setpoint, from which the agent selects the most suitable action.

However, a noteworthy challenge arises as the scale of buildings increases with the use of multiple HVAC systems. The potential divergence in the setpoints for each HVAC system introduces a substantial increase in the action space.

In large structures, this can lead to an unwieldy number of possible actions, complicating the optimization process. To address this issue, we adopted a strategic shift in the design of actions. Rather than intricately manipulating the temperature setpoints, the proposed approach focuses on the binary control of HVAC systems, specifically, toggling them on or off. This deliberate simplification effectively streamlines the action space, ensuring more manageable and efficient exploration while maintaining optimal indoor temperatures for occupant comfort and achieving energy-saving goals.

This study distinguishes itself from previous research by adopting a fundamentally different approach to DRL-based HVAC control. Whereas earlier studies have concentrated on optimizing temperature setpoints, often leading to complex and computationally intensive action spaces, our approach simplifies the control mechanism to a binary on/off decision for HVAC systems. This not only reduces computational complexity but also enhances the scalability and practicality of DRL in real-world applications. By focusing on binary control, our method offers a more straightforward yet effective solution for achieving energy efficiency and occupant comfort, setting it apart from the temperature-setpoint-driven strategies commonly explored in earlier studies.

3.2 State, action, and reward

The depiction of state and action illustrates a building environment and the conduct of agents in the environment. The state is delineated by the observations made by agents regarding the current conditions of each floor within the building. Let s_t be a set of states at time t and f be a floor in the building. The state signifies the current information for each floor in the building and each floor state is expressed as S_{ft}. Therefore, s_t can be determined using Eq. (1).

$$s_t = \{S_{1t}, \ldots, S_{Ft}\} \tag{1}$$

There are many factors that contribute to the building environment. In this study, to control temperature through HVAC systems, indoor and outdoor conditions are primarily monitored to assess the current indoor temperature. Both indoor and outdoor temperatures are key considerations. For indoor conditions, since temperature is affected by occupancy, two specific factors are considered: the total number of people inside the building and their distribution across different floors. By taking these factors into account, the system can accurately evaluate their impact on indoor temperature. For outdoor conditions, cloudiness information provided by the meteorological administration is considered, as it can affect the amount of solar radiation entering the building. Finally, time slots are also taken into account to ensure that the information is organized and analyzed effectively. S_{ft} consists of the inside and outside information of the building and is described as

$$S_{ft} = \left\{IT_{ft}, \max\left(IT_{ft} - OT_t, 0\right), \max\left(OT_t - IT_{ft}, 0\right), CN_t, DP_t, \left(SP_{ft} / MP_t\right), CT_t\right\}. \tag{2}$$

In Eq. (2), the first term, IT_{ft}, indicates the indoor temperature of floor f at time t. OT_t denotes the outdoor temperature of the building at time t. The second and third terms indicate the temperature difference between the indoor and outdoor conditions. The second term is active when the indoor temperature is lower than the outdoor temperature, and the third term is active when the indoor temperature is higher than the outdoor temperature. CN_t represents the cloud coverage at time t, which reflects the amount of sunshine. The indoor temperature is affected by the degree of sunshine. DP_t is the degree of the number of people entering the building at time t. More people entering can lead to longer door-opening times, thereby affecting the indoor temperature as the outside temperature infiltrates the door openings. SP_{ft} expresses the number of people on each floor f at time t because the presence of people affects the indoor temperature. As the range of SP_{ft} can be infinite, SP_{ft} is divided by the total number of people staying inside the building at time t, MP_t, to reduce the space. Finally, the term CT_t signifies the current position within the designated timeslot at time t. The total number of slots is determined by dividing the duration from the opening to closing of the building by the time interval at which the agent makes a decision.

In this study, the action determines the number of HVAC systems to be operated at time t, where n_f represents the number of HVAC systems installed on floor f and A_{ft} denotes the set of candidate actions. The set of candidates is expressed as $\{0, ..., n_f\}$. Therefore, the action controls the number of HVAC systems to be on at time t, and a_t can be expressed as

$$a_t = \{A_{1t}, \ldots, A_{Ft}\}. \tag{3}$$

The reward is assigned to the agent after the action and indicates the purpose of the algorithm. As mentioned previously, the main purpose of this algorithm is to maintain occupant comfort and save energy. This is not easily compatible with operating HVAC systems, leading to the use of energy, which is required to maintain occupant comfort. Equations (4) and (5) show the reward at time t on floor f and the reward formulation, respectively. As the agent selects an action to maximize the reward, its value is negative.

$$r_t = \{R_{1t}, \ldots, R_{Ft}\} \tag{4}$$

$$R_{ft} = R_{fd} + w \times A_{ft} \tag{5}$$

The reward at time t, R_{ft}, consists of two terms and is shown in Eq. (5). In the first term, UT_f denotes the upper bounds of the temperature in floor f range in which the occupants feel comfortable. R_{fd} represents the difference between IT_{ft} and UT_f. Once IT_{ft} exceeds UT_f, R_{fd} becomes $UT_f - IT_{ft}$. When IT_{ft} falls below UT, R_{fd} is 0. Equation (6) shows the value of R_{fd}.

$$R_{fd} = UT_f - IT_{ft}, \text{ if } IT_{ft} > UT_f, \text{ otherwise } 0 \tag{6}$$

In the second term of Eq. (5), the weight w is multiplied by the number of HVAC systems to be operated on floor f at time t, denoted by A_{ft}. As A_{ft} characterizes the energy consumption, the extent of HVAC system utilization is regulated by adjusting the weight w. Thus, when aiming to prioritize energy conservation, a higher value for weight w is assigned. Conversely, in situations in which energy conservation is not the primary focus and occupant comfort is more important, a lower value is selected for the weight w. Table 1 lists the notations used for DRL.

3.3 DRL algorithm for HVAC system control

In this study, the DQN algorithm was used to control the HVAC systems, as described in Algorithm 1. This algorithm runs during the operational hours of the building, defined from the opening time (1) to the closing time (T). During this period, the agent observes the state of each floor of the building at each interval i.

The action is selected by employing the ε-greedy policy or the Q-value. A random number is generated, and if the number is less than or equal to the value of ε, the action is randomly selected from the set of actions. If the number is larger than ε, the action with the maximum Q-value is selected. The value of ε gradually decreases as the algorithm is trained, increasing the likelihood that actions will be selected on the basis of the maximum Q-value over time. To generate the Q-value, the state is used as the input for the NN. The output is the Q-value corresponding to the number of operating HVAC systems. Thus, the structure of the action ranges from zero to the total number of HVAC units installed on each floor f.

Table 1
Notations for DRL.

Notation	Definition
s_t	Set of states at time t
S_{ft}	State of floor f at time t
IT_{ft}	Indoor temperature on floor f at time t
OT_t	Outdoor temperature at time t
CN_t	Cloud coverage at time t
DP_t	Degree of people entering at time t
MP_t	Total people inside building at time t
SP_{ft}	Number of people on floor f at time t
CT_t	Current position of time t within time slot
a_t	Set of actions at time t
A_{ft}	Set of candidate actions for floor f at time t
r_t	Set of rewards at time t
R_{ft}	Reward at time t of floor f
w	Weight assigned to action A_{ft}
UT_f	Upper temperature on floor f
e	Episode
T	Total time for building opening
M	Minimum training point
θ	Parameters for Q-network
θ^-	Parameters for target Q-network
i	Interval of time that agent makes decision
K	Reply memory
dr	Decay rate
ϵ	Epsilon

Algorithm 1
DQN for HVAC system control.

Input: HVAC system control problem
Output: Weights θ of the Q-network

Initialize replay memory K
Initialize Q-network with weights θ
Initialize target network Q with weights θ^{-}
for $e = 1$, E do
    for $t = 1$, T do
        if $t \% i = 0$ then
            Observe s_t
            Select a random action at with probability ε-greedy policy. Otherwise, $a_t = \max_a Q(s_t, a; \theta)$
            Execute a_t (the number of HVAC systems to be operated on floor f at time t)
                Observe s_{t+1} and assign r_t is to the agent
            Save transition (s_t, a_t, r_t, s_{t+1}) in K
            if $|Y| \geq M$ then
                Sample random minibatch (s_k, a_k, r_k, s_{k+1}) from K
                Calculate loss $\leftarrow \max_a Q(s_k, a; \theta) - \max_a Q'(s_t, a; \theta^{-})$
                Perform gradient decent regarding weight θ
                $\epsilon \leftarrow \epsilon \times dr$
            end if
        end if
    end for
    Update $\theta^{-} = \theta$
end for

Algorithm 2 presents the pseudocode for the action selection. Once an action is selected, the designated number of HVAC systems is operated for each floor. After executing the action, the next state s_{t+1} is observed and the agent receives reward r_t. The set of transitions (s_t, a_t, r_t, s_{t+1}) is stored in memory K. The memory has the maximum length, and if it is full, the old transition is removed from the memory. After the number of sets of transitions occupies more than half of the memory size, the training of the NN is initiated. Finally, the parameters of the target network are updated using the NN at the end of each episode. In the training phase, the control problem of the HVAC systems is solved and learned using the proposed algorithm. Subsequently, the trained weight of the NN is tested during the test phase. Figure 1 shows the overall framework of the DQN for the training and testing phases.

4. Case Study

The proposed control method for HVAC systems was evaluated and compared with several rule-based methods in a commercial building in South Korea. The building has five floors and five HVAC systems on each floor, all operated by a retail store. Data were collected from three floors of the building to test the proposed control method. The data were obtained from sensors and cameras installed on each floor for temperature and occupant monitoring. The data collection period spanned from March to December 2023. Figure 2 shows the framework for controlling the HVAC systems.

A simulator was designed before implementing the proposed control method in the actual HVAC systems in the commercial building. The purpose of this simulation was to test and validate the efficacy of the proposed control method in a controlled and virtual environment

Algorithm 2
Action selection

```
rand ← random():
    if rand ≤ ϵ then
        action ← random(A_ft)
    else
        action ← max Q-value(A_ft)
    end if
```

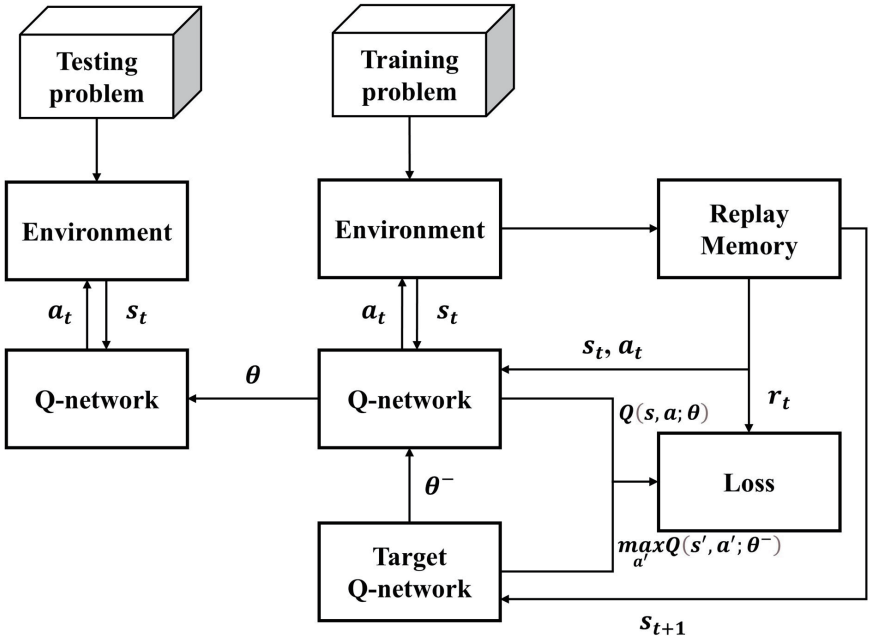

Fig. 1. DRL framework for training and testing.

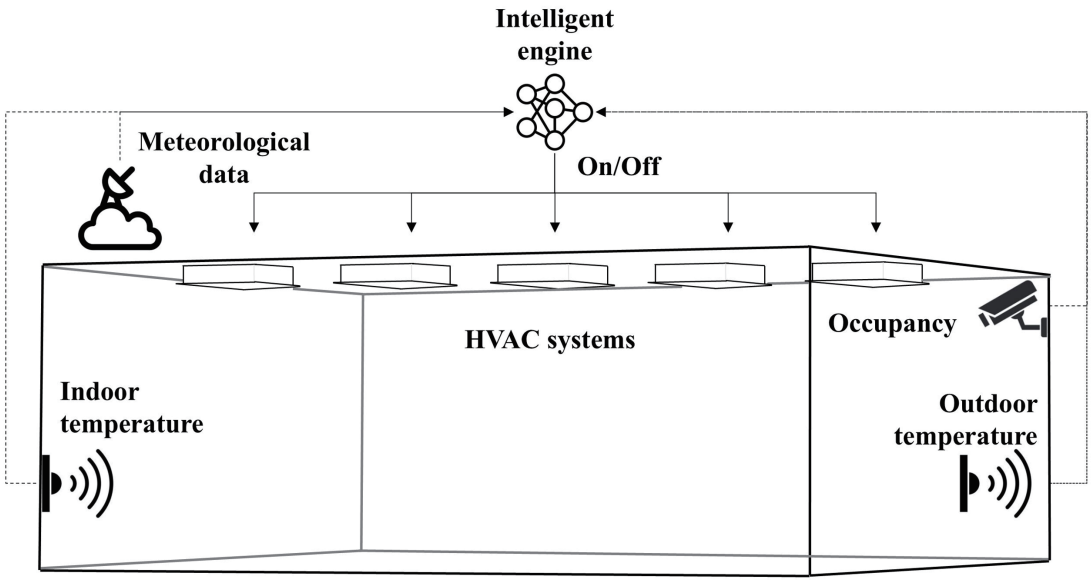

Fig. 2. HVAC system control framework.

before applying it to real-world HVAC systems. The main role of the simulator was to calculate the indoor temperature based on the operation of the HVAC systems. Diverse factors affect the variance in the indoor temperature, making it difficult to express a formulation. In this study, a deep NN (DNN) was employed to calculate the change in the indoor temperature according to the operation of the HVAC systems.

The performance of the agent was evaluated using two heuristic rules. Rule 1 states that all HVAC systems should be operated when the current temperature deviates by one degree from the upper temperature thresholds. That is, the HVAC systems should be activated if the temperature rises one degree above the upper limit. Rule 2 involves a calculation to determine the number of HVAC systems to be operated, considering the difference between the current temperature and upper temperature thresholds. The number of HVAC system operations was determined on the basis of the average ability to reduce the temperature per HVAC system.

Three key performance measurements for evaluating the efficiency and effectiveness of the proposed method were the temperature (Temp), the number of HVAC system on and off cycles (on/off), and the number of HVAC system operations during control intervals (# of on). The temperature performance measurement assessed the floor temperature, which was calculated by the simulator on the basis of the operation of the HVAC systems according to the proposed method and rules. Maintaining an optimal temperature is crucial for comfort and productivity, and this metric helps ensure that the HVAC system achieves the desired temperature levels efficiently. The measurement of the number of on and off cycles of the HVAC systems tracks how frequently the HVAC systems are turned on and off. Minimizing the number of on/off cycles is important because frequent cycling can increase the energy consumption and wear and tear on the systems. Ideally, the HVAC systems should operate smoothly with fewer on/off transitions while maintaining the desired temperature levels. The number of HVAC system operations during the control interval indicates the number of times that the HVAC systems are activated during specific control intervals. This provides insight into the frequency of HVAC system usage within a given timeframe, helping to optimize the energy consumption and operational efficiency. Minimizing unnecessary system operations without compromising temperature control is key to reducing energy costs and prolonging the equipment lifespan.

The experiment was conducted on a computer with an AMD Ryzen9 3950x processor and 64 GB of RAM. The DNN and proposed algorithm were coded in Python. The DNN predicted the indoor temperature on each floor on the basis of the operation of the HVAC systems controlled by the proposed method. The DNN employed a fully connected architecture comprising four hidden layers with 64, 128, 256, and 32 neurons. The activation function for all layers was ReLU. The input for the DNN was structured with seven parameters: the indoor temperature of floor f at time t, the outdoor temperature of the building at time t, the cloud coverage at time t, the number of people entering the building at time t, the number of people on each floor f at time t, the current position within the designated timeslot at time t, and the electricity consumption of the outdoor unit of the air conditioner. The DNN output predicted the indoor temperature on floor f at the next time step ($t + 1$).

The NN for the proposed method also had a fully connected architecture comprising three hidden layers with 64, 32, and 16 neurons. The activation function for all layers was ReLU. The

hyperparameters used in the training and testing phases are listed in Table 2. Because searching for optimal hyperparameters is difficult owing to the large search space, a random search was employed in this study to identify the best values.[28] As explained above, the agent is responsible for regulating the operation of HVAC systems in the commercial building. The agent reward is assigned on the basis of the difference between the current indoor temperature and the specified upper temperature threshold. The temperature threshold varied for each floor, as indicated in Table 3. The reason for the different temperature thresholds for each floor was that the main items were placed on the lower floor, and fewer customers tended to visit when the floor height increased.

The agent operated in a 30-min control interval, making decisions at these intervals within the operational hours of 8:00 to 22:00 when the store was open. Each day constituted one episode, with 28 transitions stored during each episode. The transitions represented (s_t, a_t, r_t, s_{t+1}) pairs in the decision-making process. During training, one of the 88 sets of real-world data was randomly selected to represent an episode. The agent learned the policy for operating the HVAC systems in a simulator on the basis of the selected actions and received rewards every 30 min. During training, if a randomly generated value was less than a specified epsilon (ϵ) value, the agent selected actions randomly. Otherwise, it selected the action with the maximum value, adhering to an exploration–exploitation strategy. The training process spanned 3000 episodes, during which the agent refined its policy by iterating through state–action pairs and the associated rewards.

The experiments for testing the proposed method were conducted over 20 days, selected randomly from the data, with each day representing a case and the operational hours ranging from 8:00 to 22:00. During these experiments, the performance of the proposed method was compared with that of two other rules across these cases. The results were then summarized by calculating the average operational hours for all floors considered, as shown in Table 4.

Table 2
Hyperparameters for DRL.

Hyperparameter	Value
Size of replay memory (Y)	20000
Optimizer	Adam
Batch size	32
Discount factor (γ)	0.95
Decaying rate (dr)	0.995
Minimum epsilon	0.01
Learning rate (lr)	0.001
Target Q-network update frequency	Every episode
Episodes (E)	3000
Q-network update frequency (C)	Every action
Minimum training points (L)	1000

Table 3
Temperature threshold for each floor.

Floor	Upper temperature (°C)
3	23
4	24
5	25

Table 4
Results of proposed method compared with other control rules.

Case	DRL			Rule 1			Rule 2		
	Temp	On/off	# of on	Temp	On/off	# of on	Temp	On/off	# of on
1	23.6	9	14.3	23.6	18.3	60	23.3	12	69
2	23.8	3	6.3	23.4	11.7	48.3	22.9	6	47
3	24	3.3	6.7	23.8	11.7	60	23.5	7.3	92.7
4	23.7	9	19	23.7	15	60	23.4	8.7	83.3
5	24.8	9.7	20	24	31.7	83.3	23.7	13.7	112
6	25.3	10.7	28	24.3	11.7	123.3	24.1	7.7	135.3
7	24.8	5	7.3	24	15	70	23.8	11	97
8	24.9	4.7	7	24.3	20	95	24	12	114.7
9	24.7	12.3	47.7	24.4	16.7	110	24	10	116
10	26.5	10.7	48	25.2	5	145	25.2	5	145
11	26.2	20	56	24.9	5	145	24.9	5.7	144
12	24.9	6	18	24	25	83.3	23.7	12.7	108
13	23.4	11	17	23.1	11.7	58.3	22.6	10	58.3
14	26	13.3	44	24.8	6.7	143.3	24.8	6	142.7
15	24.3	5.3	11	24.1	13.3	86.7	23.8	7.3	102.3
16	25	15.7	34.3	24.2	21.7	98.3	23.8	12	114.3
17	24	2.7	4.7	23.4	15	35	22.7	10	39.3
18	23.5	4.7	21.3	23.4	10	31.7	23	8	53.3
19	24.3	4.7	9.3	23.8	6.7	75	23.6	6.7	94
20	24.4	13	27.3	24	18.3	80	23.5	14	96
Mean	**24.6**	**8.7**	**22.4**	**24.0**	**14.5**	**84.6**	**23.70**	**9.3**	**98.2**

Table 5
Comparison of proposed method and other control rules for each floor.

Floor	DRL			Rule 1			Rule 2		
	Temp	On/off	# of on	Temp	On/off	# of on	Temp	On/off	# of on
3	24.6	4.6	14.5	24.2	13.0	104.25	24.0	8.05	113.8
4	24.7	4.9	15.3	24.3	13.8	101.0	24.0	8.8	111.0
5	25.2	15.9	34.5	24.2	17.5	87.5	23.8	9.1	101.7

The proposed method for controlling HVAC systems exhibited significant advantages over existing rules, particularly in terms of the on/off frequency. Better performance was achieved in 17 out of 20 cases compared with Rule 1 and in 12 out of 20 cases compared with Rule 2, indicating a notable improvement in terms of the on/off frequency. This suggests that the proposed method offers more efficient control over HVAC operations, which likely results in energy savings. Furthermore, from the perspective of the number of HVAC system operations during the control intervals (# of on), the proposed method achieved better performance in almost all cases compared with the two rules.

Table 5 shows the result for each floor between the proposed method and other control rules. Specifically, the top floor was more affected by sunlight than the other floors. Because Rule 2 operated when the temperature was above the target temperature, it showed better performance in terms of all measurements than the proposed method and the other two rules. For the remaining floors, the proposed method outperformed the other rules in terms of on/off and # of on. In terms of temperature, the control by the proposed method was worse than that of the other rules, but the difference was small.

This result showed that the operation of the HVAC systems controlled by the proposed method was better than that of the other rules. The original HVAC system was fully operated by human control. Control using the proposed method can save energy costs compared with other methods, without a significant difference in temperature.

5. Conclusion

In this study, we addressed the control problem of HVAC systems by a DRL method. The proposed approach aims to optimize energy savings while maintaining appropriate temperature levels. A case study with the real data collected by sensors was conducted to evaluate the performance of the proposed DRL-based control method, and the results were discussed.

The proposed method demonstrated superior performance compared with traditional HVAC control rules. Specifically, the DRL-based method achieved significant energy savings by reducing the frequency of HVAC system operations and minimizing the number of on/off cycles, all while maintaining the desired temperature within acceptable limits. Notably, the method showed approximately 4 to 5 times fewer HVAC system operations than conventional rules, indicating its effectiveness in balancing energy efficiency and thermal comfort.

The results suggest that DRL-based control is a promising approach to HVAC system management, outperforming conventional strategies in both energy efficiency and indoor comfort. This makes it a viable and effective solution for HVAC system control with the potential for widespread applications in various settings. Future work can focus on further optimizing and adapting the model for different building types and climates to fully leverage its capabilities.

Acknowledgments

This study was supported by the Dong-A University research fund.

References

1 A. Costa, M. M. Keane, J. I. Torrens, and E. Corry: Appl. Energy **101** (2013) 310. https://doi.org/10.1016/j.apenergy.2011.10.037
2 A. H. Buckman, M. Mayfield, and S. B. M. Beck: Smart Sustain. Built Environ. **3** (2014) 92. https://doi.org/10.1108/sasbe-01-2014-0003
3 R. Eini, L. Linkous, N. Zohrabi, and S. Abdelwahed: J. Build. Eng. **39** (2021) 102222. https://doi.org/10.1016/j.jobe.2021.102222
4 Y. Agarwal, B. Balaji, S. Dutta, R. K. Gupta, and T. Weng: Proc. 10th ACM/IEEE Int. Conf. Information Processing in Sensor Networks (2011) 246–257.
5 B. Balaji, J. Xu, A. Nwokafor, R. Gupta, and Y. Agarwal: Proc. 11th ACM Conf. Embedded Networked Sensor Systems (2013) 1–14.
6 K. Padmanabh, A. Malikarjuna, S. Sen, S. P. Katru, A. Kumar, S. K. Vuppala, and S. Paul: Proc. First ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings (2009) 37–42. https://doi.org/10.1145/1810279.1810288
7 V. L. Erickson, S. Achleitner, and A. E. Cerpa: Proc. 12th Int. Conf. Information Processing in Sensor Networks (2013) 203–216.
8 P. X. Gao and S. Keshav: Proc. Fourth Int. Conf. Future Energy Systems (2013) 237–246.
9 W. Li, C. Koo, T. Hong, J. Oh, S. H. Cha, and S. Wang: Renewable Sustainable Energy Rev. **127** (2020) 109885. https://doi.org/10.1016/j.rser.2020.109885

10 S. J. Qin and T. A. Badgwell: Control Eng. Pract. **11** (2003) 733. https://doi.org/10.1016/S0967-0661(02)00186-7
11 B. Dong and K. P. Lam: Build. Simul. **7** (2014) 89. https://dx.doi.org/10.1007/s12273-013-0142-7
12 J. Široký, F. Oldewurtel, J. Cigler, and S. Prívara: Appl. Energy **88** (2011) 3079. https://doi.org/10.1016/j.apenergy.2011.03.009
13 S. C. Bengea, A. D. Kelman, F. Borrelli, R. Taylor, and S. Narayanan: HVAC&R Res. **20** (2014) 121. https://doi.org/10.1080/10789669.2013.834781
14 M. Esrafilian-Najafabadi and F. Haghighat: Energy Build. **252** (2021) 111377. https://doi.org/10.1016/j.enbuild.2021.111377
15 M. Esrafilian-Najafabadi and F. Haghighat: Energy Build. **257** (2022) 111808. https://doi.org/10.1016/j.enbuild.2021.111808
16 Y. Peng, A. Rysanek, Z. Nagy, and A. Schlüter: Appl. Energy **211** (2018) 1343. https://doi.org/10.1016/j.apenergy.2017.12.002
17 T. Wei, Y. Wang, and Q. Zhu: Proc. 54th Annual Design Automation Conf. **2017** (2017) 1–6.
18 S. Brandi, M. S. Piscitelli, M. Martellacci, and A. Capozzoli: Energy Build. **224** (2020) 110225. https://doi.org/10.1016/j.enbuild.2020.110225
19 Z. Deng and Q. Chen: Energy Build. **238** (2021) 110860. https://doi.org/10.1016/j.enbuild.2021.110860
20 Y. Wang, K. Velswamy, and B. Huang: Processes **5** (2017) 46. http://doi.org/10.3390/pr5030046
21 K. U. Ahn and C. S. Park: Sci. Technol. Built Environ. **26** (2020) 61. https://doi.org/10.1080/23744731.2019.1680234
22 R. S. Sutton and A. G. Barto: Reinforcement Learning: An Introduction (MIT Press, 2018) 2nd ed., Chap. 6.
23 C. J. Watkins and P. Dayan: Mach. Learn. **8** (1992) 279.
24 V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, and G. Ostrovski: Nature **518** (2015) 529. https://doi.org/10.1038/nature14236
25 Y. Du, F. Li, J. Munk, K. Kurte, O. Kotevska, K. Amasyali, and H. Zandi: Electr. Power Syst. Res. **192** (2021) 106959. https://doi.org/10.1016/j.epsr.2020.106959
26 V. Hanumaiah and S. Genc: arXiv (2021). https://doi.org/10.48550/arXiv.2110.13450
27 X. Liu, M. Ren, Z. Yang, G. Yan, Y. Guo, L. Cheng, and C. Wu: Energy **259** (2022) 124857. https://doi.org/10.1016/j.energy.2022.124857
28 J. Bergstra and Y. Bengio: J. Mach. Learn. Res. **13** (2012) 281.

About the Authors

Seunghoon Lee received his Ph.D. degree in industrial engineering from Yonsei University, Seoul, Republic of Korea, in 2021. He is currently an assistant professor at Dong-A University, Busan, Republic of Korea. His research interests include artificial intelligence, optimization, and simulation in manufacturing and service industries.