

Optimizing Room Temperature Control: A Comparative Analysis of Deep Reinforcement Learning and Contextual Bayesian Optimization Methods

Zhongqi Li *

Harbin Institute of Technology (Shenzhen), Shenzhen, China

* Corresponding Author Email: 210320427@stu.hit.edu.cn

Abstract. Building temperature control presents significant modeling challenges due to the complexity of internal systems and varied usage patterns. Consequently, black box modeling has become a prevalent approach in this domain. This technique broadly encompasses deep learning and Bayesian optimization methods, each offering unique mechanisms for deriving optimal control parameters. Deep learning methods iteratively refine control parameters using extensive datasets, whereas Bayesian optimization employs Gaussian process regression and acquisition functions to pinpoint optimal solutions. This paper conducts a comparative analysis of two innovative methods within these branches: the application of Deep Reinforcement Learning (DRL) for integrated control of household temperature and electric vehicle charging, and the Primal-Dual Contextual Bayesian Optimization (PDCBO) method for adaptive thermal control in buildings. The study reveals that the DRL approach, integrated with electric vehicle charging systems, excels in responsiveness and reducing energy consumption. In contrast, the PDCBO method enhances control by allowing for the artificial definition and continuous optimization of constraints on energy usage and comfort during the operation, facilitated by algorithmic design. By synthesizing the strengths of both methods and considering electric vehicle integration and tailored adaptive control constraints, enhanced outcomes are achievable. This paper proposes that specific control strategies be developed for different building types to address diverse real-world requirements effectively.

Keywords: Comparison, DRL, PDCBO, electrical vehicles.

1. Introduction

As science and technology advance, automation systems have become increasingly integral to our society, with rising demands for sophisticated control systems. This paper focuses on the automatic control of building temperature, a critical aspect of modern building management. For any building, the internal system complexity makes direct modeling of temperature control systems nearly impractical. Challenges arise from variable electricity consumption across different floors and the additional impact of other energy-consuming facilities, such as electric vehicle charging stations, which indirectly affect the temperature control system [1]. Furthermore, daily fluctuations in external temperatures and energy use also complicate the modeling process. Consequently, direct modeling methods are seldom employed in the context of building temperature control. Instead, a black box modeling approach is more commonly adopted. This method involves selecting appropriate models based on system inputs and outputs and computing system parameters without needing to account for the system's internal structure.

Nonetheless, models developed through black box methods remain susceptible to external environmental changes. They also fall short in addressing other critical control requirements, such as occupant comfort and the transferability of the control system. Therefore, there is an imperative need to explore more advanced control methodologies. Currently, research in this area primarily revolves around two methodologies: Deep Reinforcement Learning and Bayesian Optimization [2]. These approaches offer promising pathways to enhance the adaptability and efficiency of building temperature control systems in response to complex, dynamic environmental conditions.

1.1. Deep Reinforcement Learning

Reinforcement learning methods are divided into two strategies for solving problems. One is to find better behaviors in the existing behavior space, and the other is to use statistics and dynamic programming to estimate and optimize behaviors [3]. Considering the structural differences between different buildings and the transferability of building temperature control algorithms, the first strategy is not much applicable, so strategies using statistics and dynamic programming are currently more common in this field. To implement this strategy, to build a mapping between the model's input and output, and then use building temperature control data with a sufficient sample size to train and strengthen the neural network to obtain a more accurate model [4]. If the sample size is insufficient, the model may not be able to provide more accurate output results, affecting the quality of control, which may result in failure to meet control requirements. In this case, if the sample size is not supplemented, certain constraints must be artificially imposed on the model to achieve the training effect. In addition, the daily weather also needs to be modeled, and sufficient samples are used to obtain a specific weather model through neural network training to determine the specific impact of daily weather on house temperature and improve the accuracy of the model [5].

1.2. Bayesian Optimization

Bayesian optimization is a method used to improve the objective function, which often takes a long time to complete the optimization. It consists of two main parts: one is the Bayesian statistical model used to establish the objective function model, and the other is the acquisition function that determines the next sampling point [6]. Since the purpose of Bayesian optimization is to obtain the black box function and its global optimal solution, it is also very suitable for solving the problem of building temperature control, just like the reinforcement learning method [7]. After obtaining the building temperature samples, the objective function can be obtained through the Gaussian regression process, that is, the Gaussian distribution of energy consumption and user comfort.

$$f(x_{1:k}) \sim \text{Normal}(\mu_0(x_{1:k}), \Sigma_0(x_{1:k}, x_{1:k})) \quad (1)$$

And obtain the next sampling point through the acquisition function and continuously search for the optimal value,

$$\text{EI}_n(x) := E_n [[f(x) - f_n^*]^+] \quad (2)$$

Note that in order to meet actual usage needs, it is also necessary to restrict the parameter selection of the black box model and add weather disturbances.

1.3. Research Topic and Methodology

In recent years, there has been a new research method in both directions, targeting the problem of building temperature control. The latest research in reinforcement learning takes the charging and discharging of electric vehicles into account in the temperature control system of a building, taking advantage of the time difference between the peak of electric vehicle charging and the peak of building temperature control use, and minimizing energy consumption and comfort reduction while ensuring that both needs are met [8]. The Bayesian optimization direction adopts the latest primal-dual contextual Bayesian optimization approach, which uses stricter constraints to keep energy consumption and discomfort below the set threshold in a long-term and stable manner. At the same time, the threshold can be set by the user to meet different usage requirements [9]. This paper aims to find the possibility of better automatic building temperature control by comparing and studying the energy consumption generated by the two methods and how they meet the user comfort needs [10].

2. Description of Two Theories (700)

2.1. Room Temperature Control Based on Deep Reinforcement Learning Coupled With Bidirectional Charging of Electric Vehicles

Different from the conventional reinforcement learning method applied to building temperature control, this method couples' electric vehicles charging in the parking lot into the control system. Considering that electricity prices are high during the day and low at night, and that most electric vehicles have the need to go out, the author's strategy is to reduce the charging efficiency of electric vehicles during the day and use more energy to control the temperature of the building, and the opposite at night.

The house and charging model established in this method is shown in Figure 1. The usage of electric vehicles is that they leave at 7:00 and return at 17:00. The charging fee is higher from 8:00 to 20:00, and vice versa at other times.

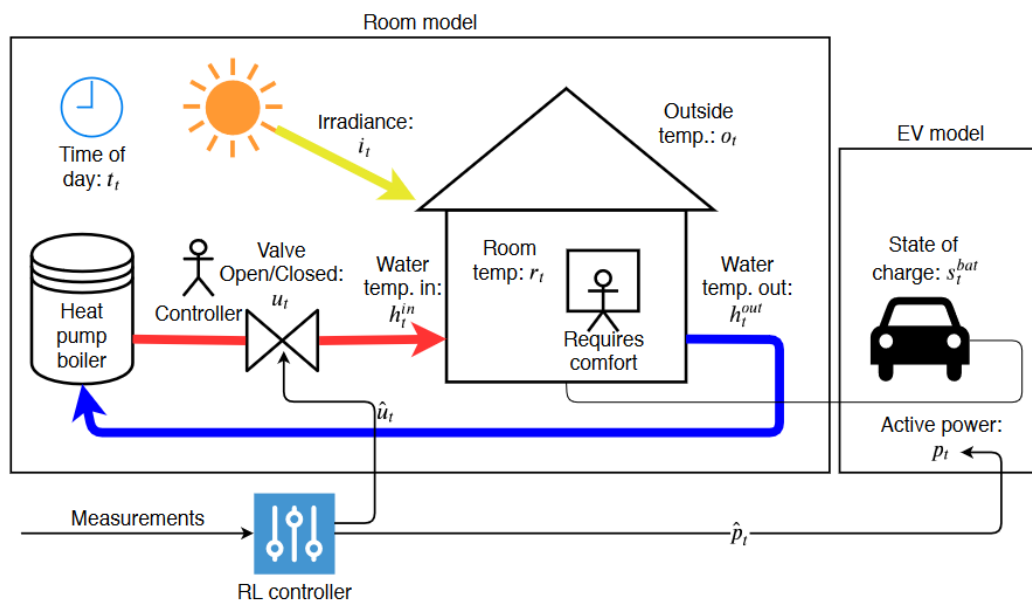


Figure 1. Overview of the room model, bidirectional EV model, and joint deep reinforcement learning controller (Photo credit: Original).

A room temperature model and a weather model were established, in which the room temperature depends on the outdoor temperature, solar radiation, pipe inlet water temperature and valve opening. At the same time, in order to prevent the room temperature model from affecting the weather output, the weather model was established separately, and the neural network was trained solely with actual weather data rather than predicted data. Based on the above room temperature and weather model, a complete building temperature model is obtained. Note that, the strategy of Recurrent Neural Network (RNN) is used to establish the temperature model, and the single-step prediction method is shown in Figure 2. By extending this method to multi-step prediction, the required neural network can be built. The electric vehicle charging and discharging model is also modeled to obtain the charging and discharging conditions of the electric vehicle, and the power of the electric vehicle is limited to the range of 20%-80% to ensure safety. In addition, the above-mentioned models are all restricted by building a deep learning environment, in order to train a better performing one when the existing data is not sufficient.

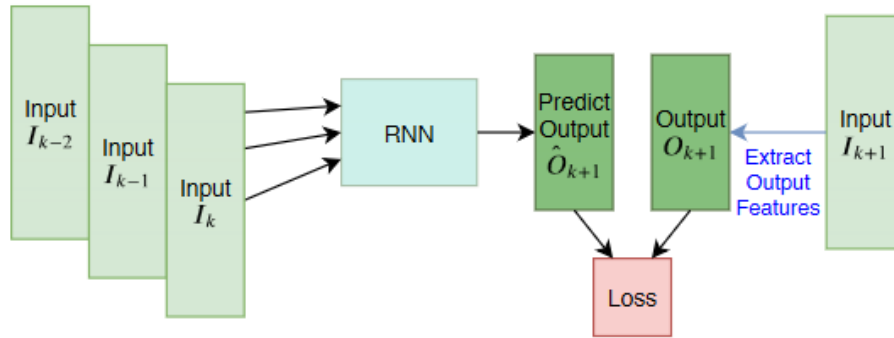


Figure 2. The simplified model uses the input from the previous three moments to predict the output at the subconscious moment and compares it with the actual output. The difference between the two values is used to train the network (Photo credit: Original).

The Deep Deterministic Policy Gradient algorithm was selected to train the Recurrent Neural Network. Since it does not contain a specific model or adopt a specific strategy, it is more suitable for completing the training of the neural network.

2.2. Room Temperature Control Based on a Primal-Dual Contextual Bayesian Optimization Approach

Similar to the above control methods, the PDCBO method also deals with various aspects of building temperature control. Apart from the electric vehicle segment, the modeling objectives of the two methods are quite similar, so a comparative analysis between them is meaningful.

In PDCBO, the controller uses proportional-integral (PI) control,

$$u(t) = K_p e(t) + K_i \int_0^t e(\tau) d\tau \quad (3)$$

The deviation signal e is the difference between the measured temperature and the set temperature.

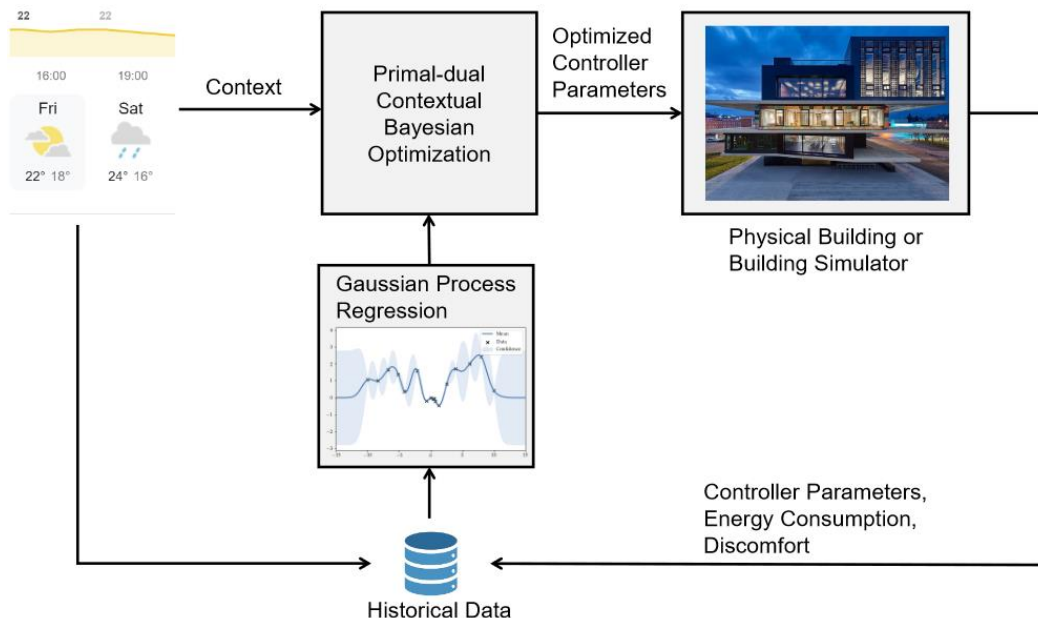


Figure 3. The general operation process of PDCBO method to control building temperature system (Photo credit: Original).

The method also takes into account the difference between the operation of the temperature control system during the day and at night, entering the night mode at 6 pm and setting the target temperature at night to 22.5°. The target temperature is used to quantify the discomfort, that is, the integration of the absolute value of the difference between the actual temperature and the set temperature. This way of setting discomfort is consistent with that in DRL control. In addition, points are also earned

for system energy consumption, with the object of integration being the power of the control system. The average daily ambient temperature and the average daily solar radiation are considered as the most important environmental factors and used as the input of the algorithm context variables, denoted by zn . Since the true values of the average daily ambient temperature and average daily solar radiation on the second day cannot be obtained in actual situations, the predicted values of the weather forecast are used as context variables, and only the predicted values on the second day are used to ensure their accuracy as much as possible. Figure 3 is the general operation flow of the PDCBO method.

As mentioned above, the daily weather forecast values are passed into the PDCBO link, and the temperature control system parameters are obtained by PDCBO. After the control of the day is completed, the control parameters, energy consumption, discomfort and weather forecast values of the control system are input into the relevant algorithm as the posterior values of the Gaussian process, and the constraints on discomfort and energy consumption and the Gaussian process parameters are updated and put into use on the next day. Due to the existence of the update algorithm, the system's initial energy consumption and discomfort constraints can be set autonomously and adjusted according to the operating conditions.

3. Performance Comparison

3.1. System Performance of DRL Strategy

The comparison between the battery charging and discharging model and the actual situation as well as the average absolute error and maximum absolute error in the prediction process are shown in Figure 4.

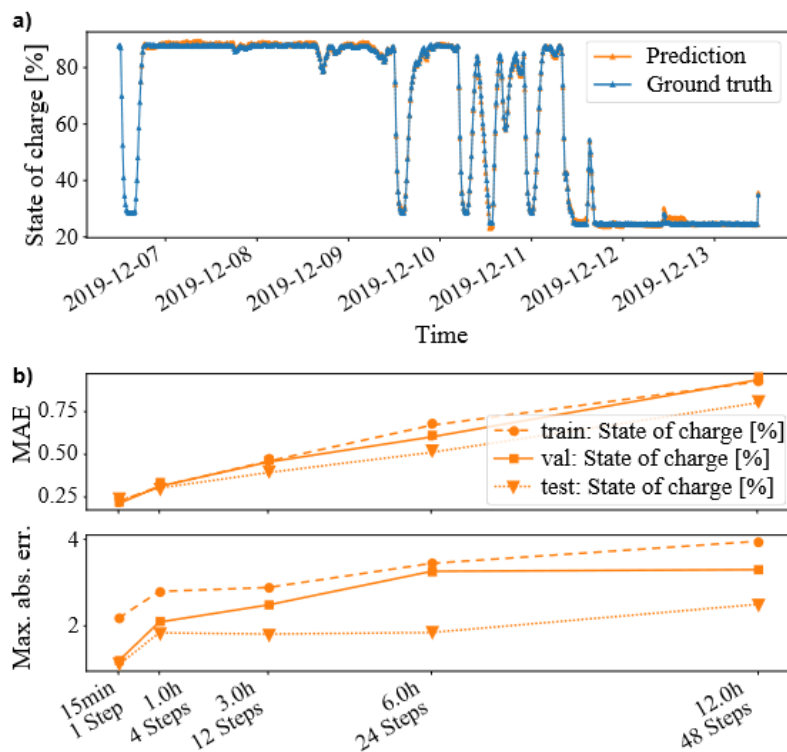


Figure 4. Prediction performance of the EV battery model. a) A sample week from the training set. b) Quantitative evaluation (Photo credit: Original).

It can be seen from the figure that the predicted value of the charge and discharge model is in good agreement with the actual situation. The MAE and maximum absolute error are also kept within a small range.

The simulation of the weather model is shown in Figure 5.

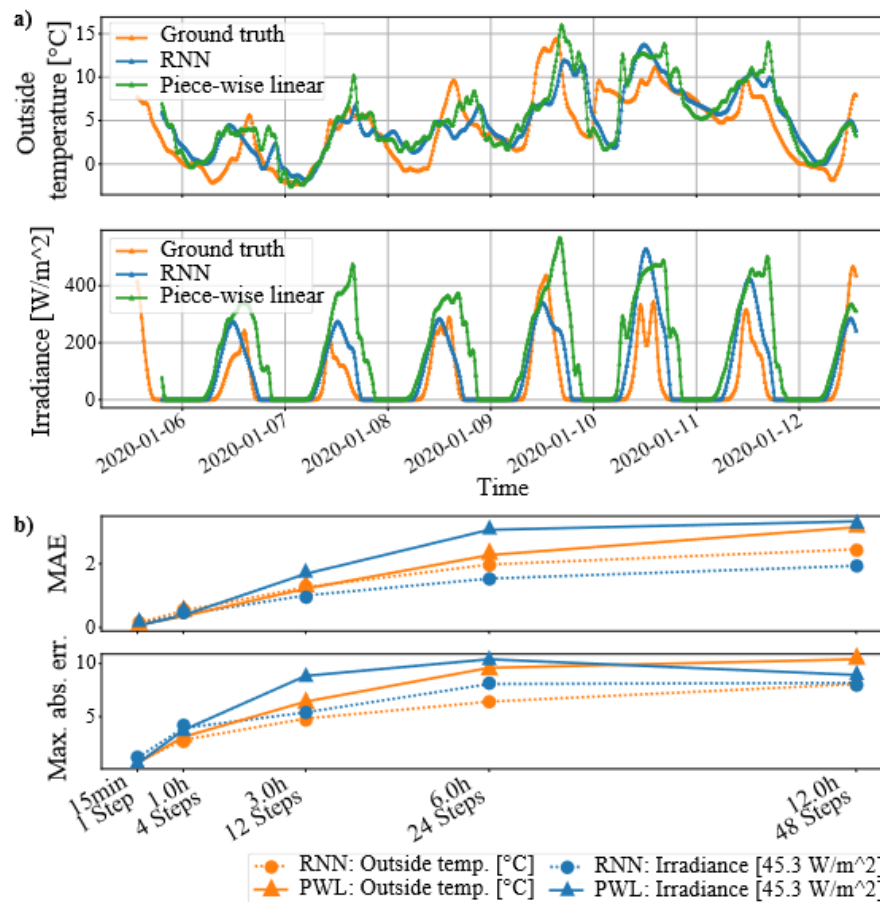


Figure 5. Weather models: a) Prediction over several days example. b) Quantitative comparison (Photo credit: Original).

As a comparison, a piecewise linear model is also introduced. It can be seen that although the tracking error performance of the two methods is not much different, the tracking ability and change smoothness of the piecewise linear model are far inferior to those of the RNN model.

Figure 6 shows the simulation data of controlling the temperature of a house using the DDPG strategy coupled with the electric vehicle charging and discharging method.

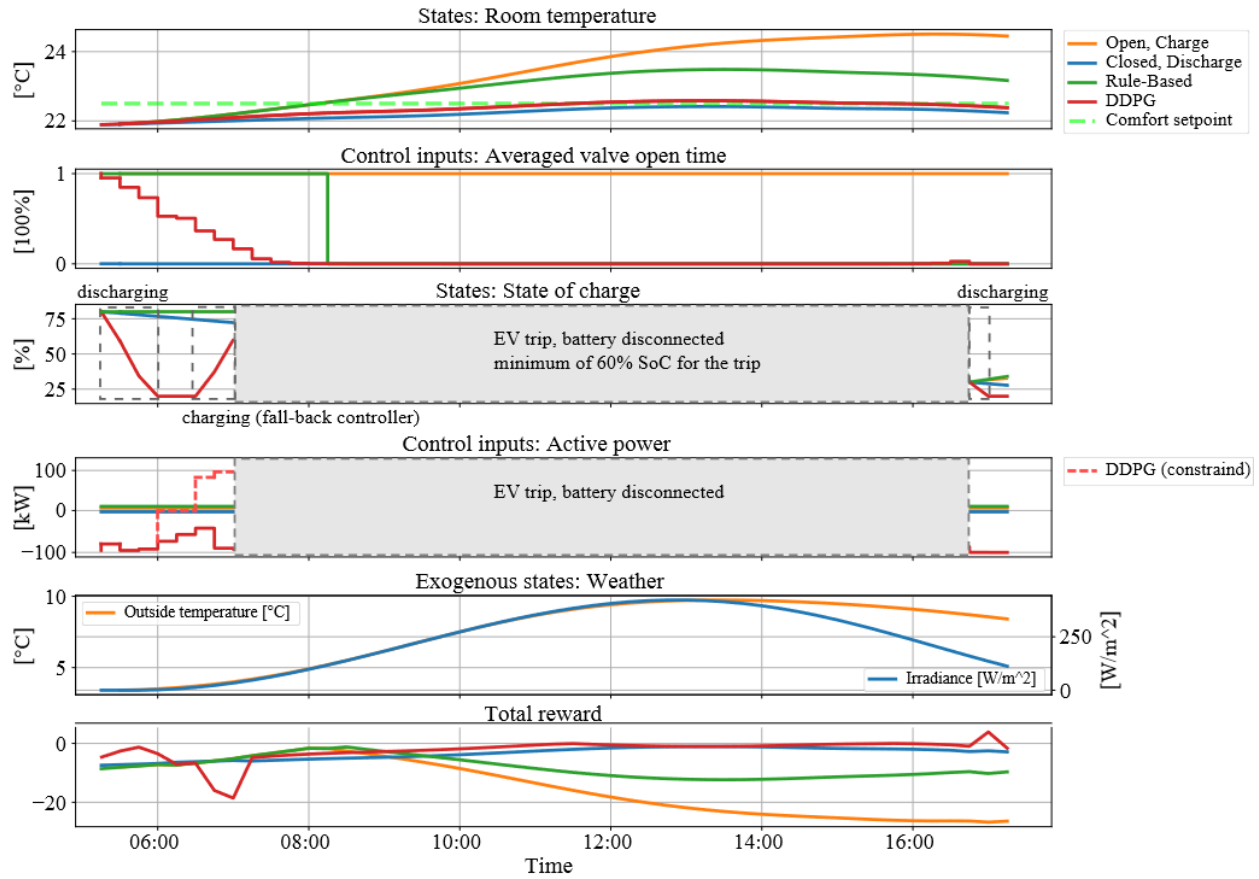


Figure 6. Joint room heating control and EV charging. Evaluation of control agents in simulation - example. Inflow water temperature: 29.4°C (Photo credit: Original).

The figure compares DDPG with three other control strategies. Among them, "Open, Charge": keep the water inlet valve fully open and continue to charge the electric vehicle at maximum power when it is not fully charged; "Closed, Discharge": keep the water inlet valve closed and continue to discharge the electric vehicle at maximum power; "Rule-Based": control the valve with the "Bang-Bang" strategy, that is, fully open the valve when it is lower than the set temperature, and close the valve when it is higher than the set temperature, while keeping the battery charged at maximum power. The DDPG strategy gradually adjusts the valve opening according to the difference between the current temperature and the set temperature. Furthermore, when more energy is needed for temperature control, energy is called from the electric vehicles to quickly achieve the goal, and then they are charged to an allowable range before the set car travel time.

Figures 7 and 8 show the data obtained when the control system was put into actual application.

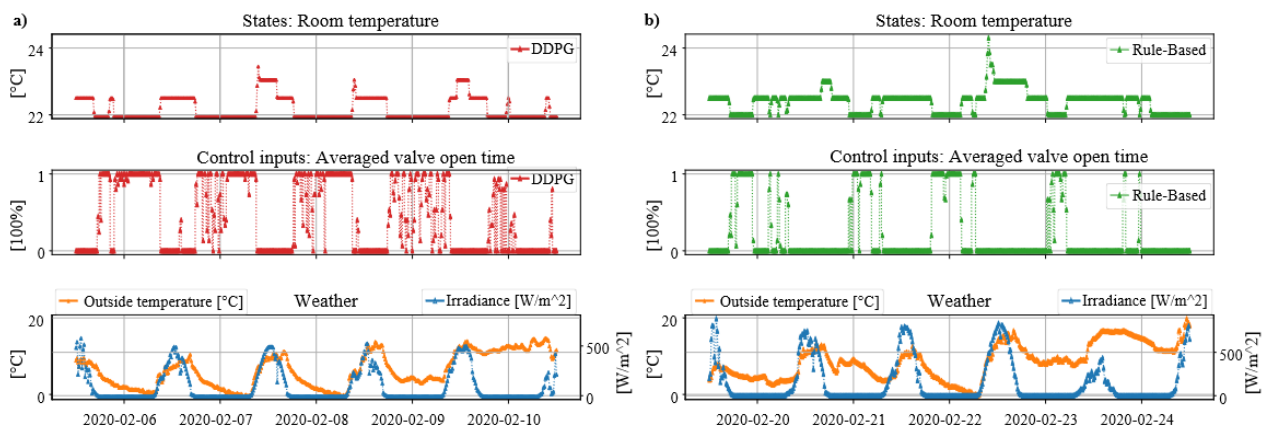


Figure 7. The return data obtained after actual application. a) DDPG strategy. b) Rule-Based strategy (Photo credit: Original).

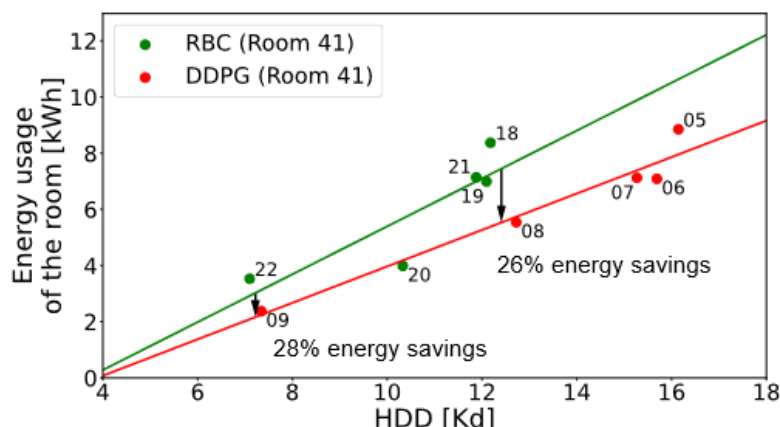


Figure 8. Comparison of energy consumption of RBC and DDPG strategies under different HDDs during the sampling time (Photo credit: Original).

This time, only the DDPG and RBC control strategies were put into formal operation. It can be seen that although the house temperature is maintained at 22.5°C for a shorter time under the DDPG strategy than under the RBC strategy, the valve opening time when DDPG is running, which can represent energy consumption to a certain extent, is less than that of the RBC strategy. And from Figure 8 can see that the operating environment temperature of DDPG is lower than the control environment of RBC. For comparison, HDD (Heating Degree Days) is introduced to indicate how close the temperature is to 18°C in a day. Therefore, by comparing the two strategies after normalization, it can be found that the average energy consumption of the DDPG strategy is reduced by 25%-30% compared to the RBC strategy.

3.2. System Performance of PDCBO Strategy

In the PDCBO method, because the constraints of the control system on energy consumption and discomfort can be set artificially, two control strategies are adopted, namely, limiting discomfort to minimize energy consumption and limiting energy consumption to minimize discomfort.

Figure 9 shows the dynamic performance of the system at different thresholds under the strategy of limiting discomfort and minimizing energy consumption. Note that 1Kh represents a deviation from the target temperature of 1 Kelvin for 1 hour, and the daytime energy consumption is doubled to correspond to the different electricity prices during the day and at night, which is similar to the distinction between electricity prices at different times of the day in the DRL control strategy. In addition, the strategies SafeOPT, which is more conservative and focuses more on reducing energy consumption, and CEI, which is more aggressive and focuses on reducing discomfort, are used for comparison.

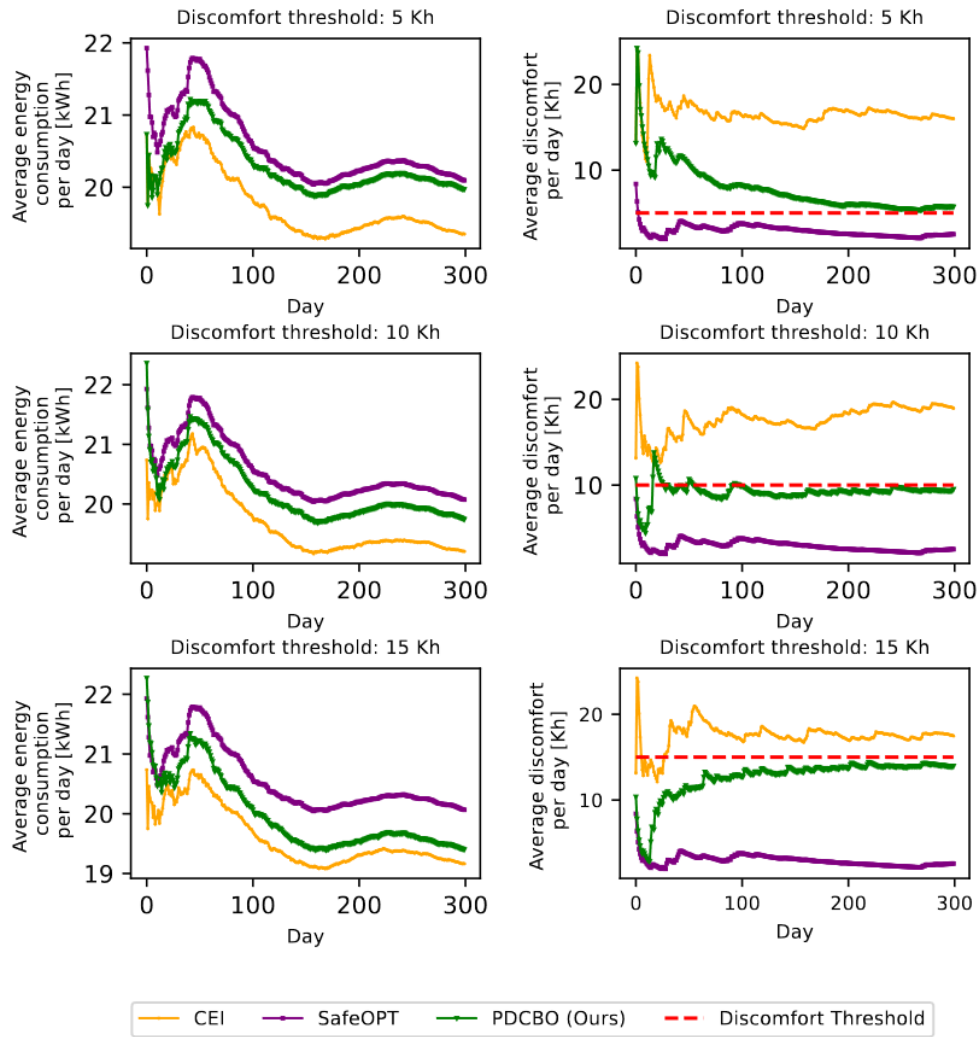


Figure 9. Daily discomfort and average energy consumption. Discomfort thresholds are 5Kh, 10Kh, and 15Kh (Photo credit: Original).

In addition, the algorithm adopted by PDCBO can modify the constraints during the control process. Figure 10 shows the system dynamics caused by changing the discomfort threshold at different stages over 300 days.

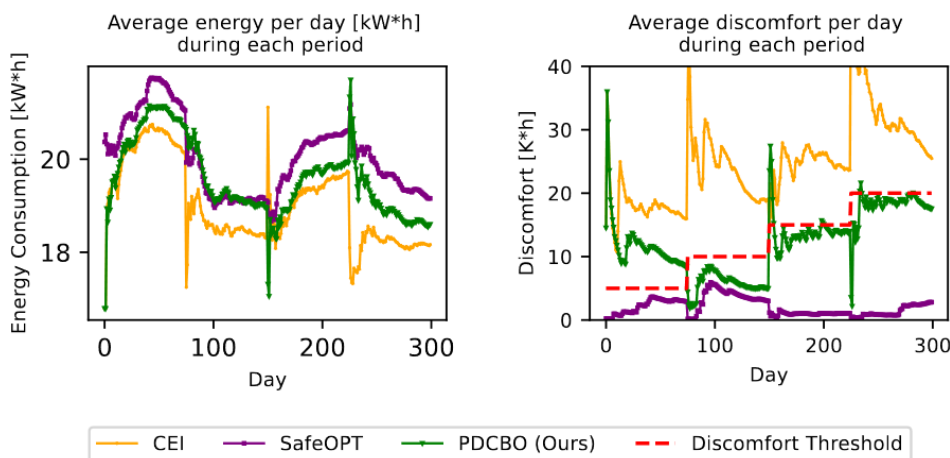


Figure 10. Dynamic performance of the system over 300 days. The discomfort threshold was changed at 75, 150, and 225 days (Photo credit: Original).

Figure 11 shows the system energy consumption and discomfort under the strategy of limiting energy consumption and minimizing discomfort. Three different energy consumption threshold limits

are also adopted. Compared with the SafeOPT strategy, the discomfort was reduced by 46%, 53%, and 63%, respectively.

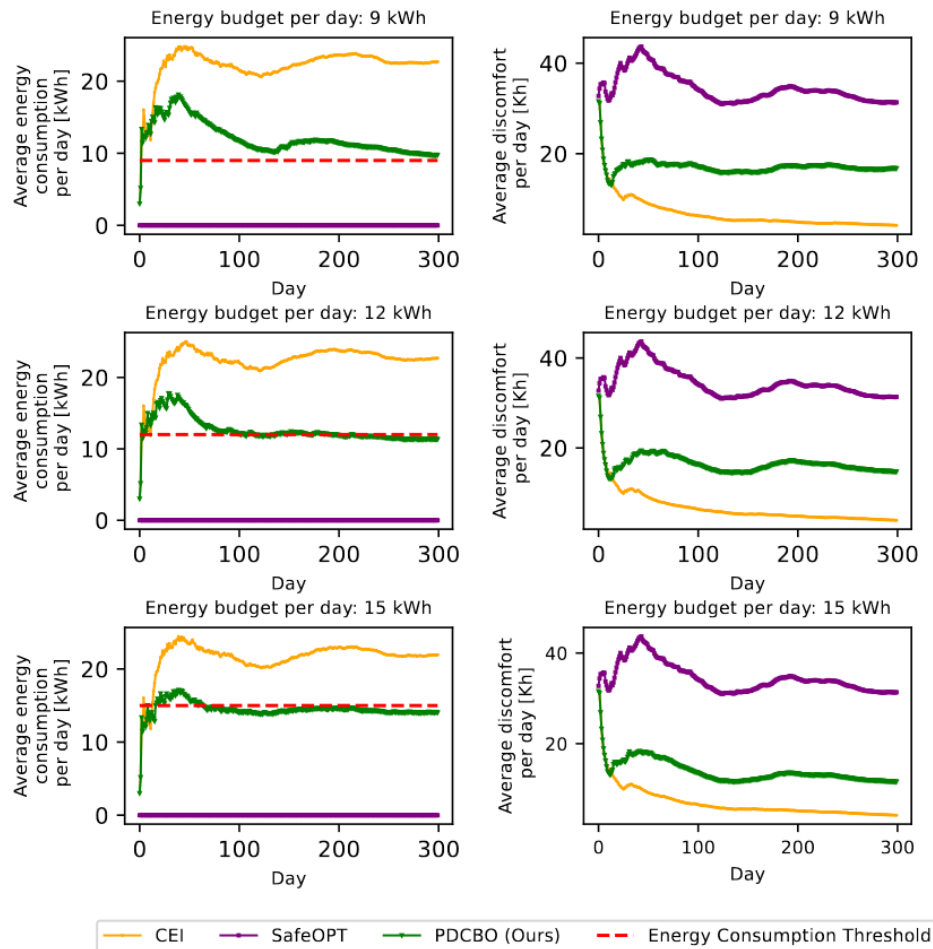


Figure 11. Daily discomfort and average energy consumption. Energy consumption thresholds are 9kWh, 12kWh, and 15kWh (Photo credit: Original).

3.3. Difference and Reasons

According to the simulation data of the DRL control method, since the goal is to limit energy consumption as much as possible, the valve opening is completely closed when there is still a distance between the indoor temperature and the target temperature. It takes a long time for the target temperature to be reached through heating by the weather, so it can be seen that there is a large accumulation of discomfort in the early stage of control. The data obtained from actual operation also show that under the DRL control strategy, the room temperature is not at the target temperature for a long time, but the valve is open for a shorter time than RBC and most of the time the valve is only partially open. It can also be seen in the PDCBO control method that when the constraint threshold is small, the algorithm violates the constraint to a certain extent so that another unconstrained performance parameter remains within a reasonable range. In fact, when the constraints are small, appropriate excess will not have much physical impact on people in the building. It can be seen in Figure 7 that the temperature only varies within the range of 22-23°C. In this case, the two control methods agree.

On the other hand, when the constraint threshold is large, exceeding the constraint will significantly increase discomfort or energy consumption even at the same ratio. Therefore, under PDCBO control, the constraint items are kept consistent with the threshold as much as possible, and because the threshold is large, the allowable range is not too strict, and the unconstrained item is also kept at a moderate value. As for the DRL strategy, although the operating conditions with a larger constraint threshold are not given, it can be predicted that in this case the algorithm will still spend

more time letting the weather adjust the system, thereby achieving the goal of sacrificing discomfort and reducing energy consumption.

Apart from the above comparable aspects, the two methods each have their own unique features. In the DRL control method, due to the introduction of the electric vehicle model, the building temperature can be controlled at a faster speed without affecting the use of electric vehicles. In addition to energy consumption and discomfort, this strategy also reduces the cost of charging, bringing better environmental and economic benefits. In the PDCBO control method, since the constraints can be modified manually, the needs of different users can be met, and even modifications can be made midway during the control process to cope with changes in the users of the building or house, which is more beneficial to them.

3.4. Making Suggestions

After comparing the two excellent control methods horizontally and vertically, it is necessary to extract their advantages and apply them to future building temperature control. As for the performance of the control system itself, the PDCBO method is better and can complete the control task well under different constraints. The idea of taking electric vehicles into consideration in the control system in the DRL method is very good, especially in the current market situation where electric vehicles are being used more and more, and the application prospects of this method are great. At the same time, for different application environments, such as the opposite usage of cars in office buildings and residential buildings, different models can be constructed to achieve the control objectives.

4. Challenges

Although the two control methods have similar modeling for each link, the actual operation conditions are different. The DRL method focuses more on observing the changes in energy consumption and discomfort within a day, and uses the RBC method for comparison. The PDCBO method, on the other hand, focuses more on the long-term changes in daily average indicators, and is compared with the strategies of SafeOPT and CEI. Therefore, it is difficult to conduct a quantitative analysis of the two methods. However, qualitative analysis is entirely possible, which allows us to clearly understand the advantages and disadvantages of each method and hope to leverage these advantages in future control system designs.

5. Conclusion

This paper evaluates two methodologies for automating building temperature control: one leveraging deep reinforcement learning (DRL) for managing house temperature and electric vehicle charging, and another employing primal-dual contextual Bayesian optimization (PDCBO) for thermal control with adaptively adjusted constraints. Analysis reveals that while the DRL approach struggles to maintain the system temperature at a set target for extended periods due to suboptimal constraint selection, thus enhancing energy efficiency, it excels when integrating electric vehicle charging and discharging into the building temperature control system. The DRL method employs a straightforward bifurcated strategy to distinguish between user needs for temperature control and electric vehicle charging. Conversely, the PDCBO method demonstrates superior performance within the control system. By utilizing both historical and recent data to update Gaussian process parameters each cycle, it enables real-time adjustment of system constraints based on current usage, thus facilitating constraint customization and self-adaptation. This method allows for mid-operation modifications to accommodate varying user demands, offering potential enhancements in user comfort and utility for architectural design, particularly if different performance constraints are tailored for various building types, such as residential homes, office buildings, and shopping centers. Currently, few control strategies consider electric vehicle charging within temperature management

frameworks. Thus, incorporating this module could significantly refine existing approaches. For the PDCBO method in particular, integrating electric vehicle charging and discharging could markedly improve system responsiveness and reduce energy consumption. This integration promises substantial advancements in building environmental control technologies.

References

- [1] Svetozarevic, B., Baumann, C., Muntwiler, S., Di Natale, L., Zeilinger, M. N., & Heer, P. (2022). Data-driven control of room temperature and bidirectional EV charging using deep reinforcement learning: Simulations and experiments. *Applied Energy*, 307, 118127.
- [2] Xu, W., Svetozarevic, B., Di Natale, L., Heer, P., & Jones, C. N. (2024). Data-driven adaptive building thermal controller tuning with constraints: A primal-dual contextual Bayesian optimization approach. *Applied Energy*, 358, 122493.
- [3] Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement Learning: A Survey. *Journal of Artificial Intelligence Research*, 4, 237 - 285.
- [4] Frazier, P. I. (2018). A Tutorial on Bayesian Optimization. *arXiv preprint arXiv: 1807.02811*.
- [5] Fiducioso, M., et al. (2019). Safe Contextual Bayesian Optimization for Sustainable Room Temperature PID Control Tuning. *arXiv preprint arXiv: 1906. 12086*.
- [6] Zhu, X., Zhao, Z., Wei, X., & Others. (2021). Action Recognition Method Based on Wavelet Transform and Neural Network in Wireless Network. In 2021 5th International Conference on Digital Signal Processing (pp. 60-65). [Conference Location].
- [7] De Blasi, S., et al. (2024). Safe Contextual Bayesian Optimization Integrated in Industrial Control for Self-Learning Machines. *Journal of Intelligent Manufacturing*, 35 (2), 885 - 903.
- [8] Char, I., et al. (2019). Offline Contextual Bayesian Optimization. In *Advances in Neural Information Processing Systems* 32.
- [9] Chakrabarty, A., et al. (2021). Accelerating Self-Optimization Control of Refrigerant Cycles with Bayesian Optimization and Adaptive Moment Estimation. *Applied Thermal Engineering*, 197, 117335.
- [10] Kazmi, H., et al. (2018). Gigawatt-Hour Scale Savings on a Budget of Zero: Deep Reinforcement Learning Based Optimal Control of Hot Water Systems. *Energy*, 144, 159 - 168.