Human in the Loop Learning for Energy Savings in Centralized Air-Conditioned Public Spaces

Anonymous Author(s)
Affiliation
Address
email

Abstract

Reducing global energy consumption is an urgent step towards slowing the pace of global warming and climate change. Many opportunities to do so lie in the building sector, which is a major energy consumer, particularly Heating, Ventilation and Air-Conditioning (HVAC) systems. Centralized air conditioning in large communal spaces (e.g. malls, offices, cinemas, libraries) often aims to cool spaces to very low temperatures, leading to significant energy consumption that may not even be necessary for users' comfort. Therefore, we propose a human-in-the-loop aided reinforcement learning (RL) framework, *HLAC RL*, to optimize energy savings in public buildings while maintaining user comfort. By collecting user input on the current environment temperature (e.g., thermal perception feedback) through an application, we can adaptively aggregate and incorporate user needs into the RL algorithm, with the dual goals of ensuring user comfort and reducing energy consumption. This user feedback thus serves as a "communal thermostat" that can monitor and optimize HVAC actions given variability across users' comfort levels, as well as changes to the external temperature, the set of users in the building, and their comfort levels over time. HLAC RL outperforms a Set-point Control baseline during low-activity periods by identifying the specific needs of smaller user groups through more personalized temperature decisions, which can markedly differ from the generic optimal temperature solution for the average user.

1 Introduction

According to the US Department of Energy [1], reducing energy usage is key in the fight against climate change because it will reduce the burning of fossil fuels in traditional power plants and, thus, reduce the release of greenhouse gases (GHG) and air pollution. In view of the need for fast-acting solutions, we propose that one key area of focus can be the development of better temperature control solutions for large-scale cooling systems in buildings. The building sector is one of the major energy consumers in the world, consuming around 36% of the total global energy consumption [2], with Heating, Ventilation and Air-Conditioning (HVAC) systems being the largest contributors [3]. In tropical regions particularly, effective cooling systems are essential to ensure the thermal comfort of building occupants. In public buildings, it is estimated that cooling accounts for more than 50% of energy usage [3]. Temperature control solutions can then significantly reduce energy usage and can also be incorporated into existing cooling systems easily and relatively cheaply, thus providing an avenue for quick climate action given the urgency of dealing with global warming.

Submitted to Tackling Climate Change with Machine Learning: workshop at NeurIPS 2023.

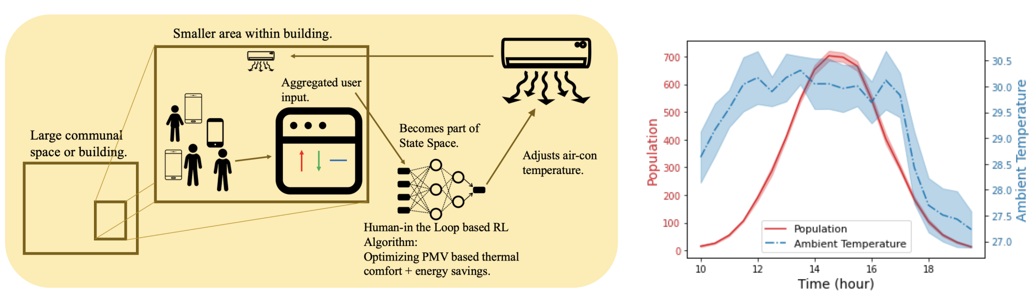
In reducing energy consumption, it is also important to preserve user comfort levels. We consider a highly occupied public space, such as a public mall, a cinema, a library, or a communal office space. Such buildings usually have a centralized air-conditioning system. Unlike individual households or offices with thermostats and temperature switches, public spaces often do not allow users to control the temperature, partly due to the presence of many users with potentially conflicting preferences. Nevertheless, users may find the temperature too cold. At the same time, increasing the temperature could also bring about energy savings.

Contribution: We propose a novel human-in-the-loop reinforcement learning framework for air-condition temperature control (HLAC RL), which aggregates real-time user feedback, and learns how to adjust the temperature in a way which balances optimizing energy savings with user-comfort. Our system obtains *crowd-sourced* user feedback. Users can input their personalized preferences of the current temperature into an app to indicate whether they prefer a higher or lower temperature. User input is exactly mapped to one of 9 thermal sensation modes, and thermal comfort is predicted, via a function adhering to the standards proposed by the *American Society of Heating, Refrigerating and Air-Conditioning Engineers* (ASHRAE) [4, 5]. Reinforcement learning is used because explicitly formulating a policy is not possible, due to the variability in human sensations and the dynamic environment.

Our framework is novel as a) it operates in an online setting, incorporating *real-time user feedback* instead of relying on functions learned from historical data [6, 7, 8, 9, 10], for *temperature control in public and communal spaces*, and b) it seeks to determine the optimal temperature for energy savings while maintaining user comfort. We will release the simulator for other researchers, industry partners and the public to use. The different components of the simulator can be modified, and customized for the specific building/public space setting.

2 Human in the Loop Learning Solution

We design a human in the loop aided RL framework HLAC RL (Fig. 1), to balance optimizing user comfort and energy savings, as an alternative to centralized aircon systems in large communal spaces. Firstly, we model a public space, with the user arrival and hourly temperature distributions following the time of day. Details are in Section 3. At every pre-defined interval, *user input* (their thermal perception feedback) on the current indoor temperature is received through an app: Users vote whether they prefer an increase in temperature, or a decrease in temperature. This feedback is *aggregated, and incorporated* into the RL algorithm, serving as a communal thermostat.



(a) The Human in the Loop Temperature Control (HLAC RL) Pipeline.

(b) Daily Simulation of population and outdoor temperature

Figure 1: Details of our Human in the Loop Learning Setup, and the Environment Simulation.

We aim to maximize the following reward function:

$$Reward(t) = w_c \frac{\sum_{i=0}^{|Population(t)|} UserComfort(user_i, temp_t)}{|Population(t)|} - w_e EnergyUsage(temp_t), \quad (1)$$

which is a weighted sum of user thermal comfort (*UserComfort*(t)) and energy savings (*EnergyUsage*(t)), with w_c and w_e representing the respective weights. By adjusting these weights, the algorithm can shift its focus, emphasizing one aspect over the other.

UserComfort: A data-driven method is used to obtain the thermal comfort function. The *predicted mean vote* (PMV) is a model used in the ASHRAE-55 standard [5], which we use as the base function for modelling individual thermal sensation in this work. Comfort levels are on a scale from -3 to +3, where -3 represents very cold, and +3 represents very hot. The PMV calculation depends on 6 factors: the metabolic rate (met), clothing insulation (clo), dry bulb air temperature, mean radiant temperature, relative humidity, and relative air velocity [11]. We use the ASHRAE Global Thermal Comfort Database [12] to model 9 thermal comfort modes of users (See Appendix B for further details). Users are sampled from these 9 thermal comfort modes and their input is generated accordingly, then we use that input to (with 100% accuracy) infer which mode corresponds to which user. Based on the current indoor temperature and these 9 modes of users, the PMV score is predicted. The predicted PMV score is further transformed into a penalty (i.e. cost) value by a function. Per ASHRAE-55's recommended thermal limit, the function places a higher penalty for scores outside the range of $[-0.5, 0.5]$ [4]. The PMV and penalty for one of the nine modes is shown in Fig. 2a.

$$\textit{Penalty} = -0.1 \times |\mathrm{PMV}| \times (\mathrm{PMV}^4 + 1) \tag{2}$$

EnergyUsage: the function is modelled based on the heat transfer equation [13]:

$$\textit{EnergyUsage} = \frac{mc_a \Delta T}{\textit{EER}}, \tag{3}$$

where m is the mass of the room, c_a is the heat capacity of dry air, ΔT is the temperature difference between the outdoor (ambient) temperature and the indoor (air-con controlled) temperature, and *EER* is the ratio of the cooling capacity to the power input.

To optimize the reward function, our **state space** is an amalgamation of the *human input* [number of votes for "temperature increase", number of votes for "temperature decrease"], along with the variables [current time, outdoor temperature, indoor temperature]. We input this state space into a deep Q-learning algorithm [14]. The output of the neural network is the **action** taken: the temperature of the air-con. For safety reasons, the actions (temperature set) range between $20 - 28°C$, with intervals of $0.2°C$. In Q-learning, the Q-function $Q(s, a)$ represents the value (sum of expected discounted reward) of taking action a at state s, and the algorithm's aim is to iterate until convergence at the true Q-value.

$$Q(s, a) \leftarrow (1 - \alpha)Q(s, a) + \alpha(Reward(t) + \gamma \max_a Q(s', a)), \tag{4}$$

where α is the learning rate, and γ is the discount factor. In deep Q-learning, the neural network parameterizes the Q-function. The loss function which the neural network optimizes is

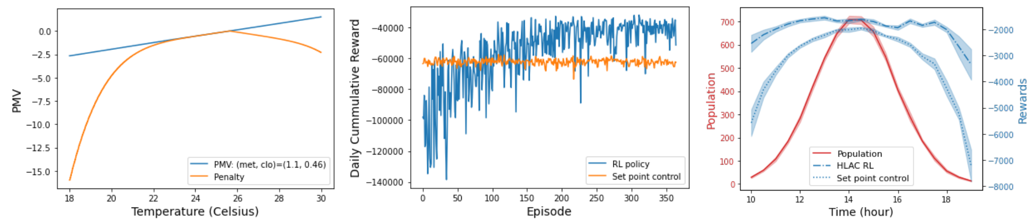
$$L = [Reward + \gamma \max_{a'} Q(s', a'; \theta') - Q(s, a; \theta)]^2. \tag{5}$$

3 Simulation Setup and Results

Daily and Hourly Simulation of the public space: An open space is modelled, to mimic a large shopping mall with 4 floors covering an area of 4890 m^2 and accommodating a total of 6000 consumers in a day [15]. The space operates from 10 a.m. to 8 p.m. The outdoor environment is simulated to match the summer conditions of a tropical climate, following data on the average hourly temperature variations in Singapore from 1896-2016 [16]. The arrivals of consumers are modeled using the following function:

$$\textit{Population}(t) = \textit{Population}(t-1) + \textit{In}(t) + \textit{Out}(t) \tag{6}$$

The *Population* function refers to consumers who are in the space; *In* refers to those who arrive; *Out* refers to those who will leave the space. The *Population* at each timeslot in a day follows a normal distribution, determining which users leave and which stay. The simulation is conducted with a granularity of 30 minutes. All parameters can be easily customized. Fig. 1a shows the simulation system. Fig. 1b shows the hourly variations of indoor population and outdoor temperature after running 10 times, with a confidence interval of 95%.



(a) PMV and penalty

(b) Convergence graph

(c) Hourly comparison: HLAC RL vs. Set point policy

Figure 2: Penalty function graph; training process; and hourly performance

We describe the simulation parameter values in Appendix A. For the baseline policy, we use the *Set-point Control* (fixed policy) technique of VRF systems [17], as VRF systems are often used in large buildings. Just like [17], we use a constant temperature of $26°C$, as ASHRAE 55-2017 [5] recommends a summertime thermal comfort zone of $23 - 26°C$.

Fig. 2b shows how the learned reward value changes during the training process of HLAC RL . Each episode represents a day-long simulation, with rewards calculated at every half hour interval. The daily cumulative reward is the sum of these rewards. HLAC RL outperforms the set point policy (fixed temperature of $26°C$) after learning for 100 episodes and converges after 300 episodes. Fig. 2c shows the hourly variations of the population and the reward obtained using HLAC RL and the set point policy after running each 10 times, with a confidence interval of 95%. The performances fluctuate throughout the day, largely due to the varying user activity levels. During low-activity periods (morning or evening), the HLAC RL can effectively cater to the specific needs of the smaller group by employing personalized temperature decisions, outperforming Set-point Control. With increased users, the HLAC RL transitions to a generic decision, performing similarly to Set-point Control for the larger, diverse group.

4 Pathway to Impact and Future Work

Pathway to impact: Given that thermostats and temperature switches in personal units in buildings are used to adjust temperatures, it is possible to integrate a crowdsourced human-in-the-loop feedback mechanism as an analogous thermostat for temperature control in large public spaces. The neural network used for computations is lightweight. As air-conditioning in buildings is a major consumer of energy [3], developing this human in the loop temperature control solution provides an avenue for quick climate action, through the reduction of GHG emissions. We demonstrated that HLAC RL excels in making personalized temperature control for smaller groups, outperforming the set-point control policy there. Its impact can be improved by implementing it across multiple subsections within large spaces.

Future work: We proposed a novel human-in-the-loop temperature control framework. We will release this building simulator, with HLAC RL algorithm integrated in, for open use. In future work, to further improve the usability of this novel framework, a) we will statistically analyze and characterize to what extent we can detect attacks in the form of intentionally noisy user input and b) explore a pure model free approach of aggregating user input, complementing the 9 thermal modes in this work. Furthermore, the energy usage and thermal comfort prediction may be modified according to the complexity desired. We could introduce a hybrid RL architecture with separate models for the two aspects, which are then combined for joint learning.

References

[1] US Department of Energy, "Energy efficiency: Buildings and industry," https://www.energy.gov/eere/energy-efficiency-buildings-and-industry, accessed: 2023-09-30.

[2] M. Santamouris and K. Vasilakopoulou, "Present and future energy consumption of buildings: Challenges and opportunities towards decarbonisation," *e-Prime-Advances in Electrical Engineering, Electronics and Energy*, vol. 1, p. 100002, 2021.

[3] A. R. Katili, R. Boukhanouf, and R. Wilson, "Space cooling in buildings in hot and humid climates—a review of the effect of humidity on the applicability of existing cooling techniques," in *14th International Conference on Sustainable Energy Technologies â€"SET*, 2015.

[4] A. Standard, "Thermal environmental conditions for human occupancy," *ANSI/ASHRAE, 55*, vol. 5, 1992.

[5] A. ASHRAE, "Standard 55-2017," *Thermal environmental conditions for human occupancy*, 2017.

[6] Y. Huang, A. Khajepour, H. Ding, F. Bagheri, and M. Bahrami, "An energy-saving set-point optimizer with a sliding mode controller for automotive air-conditioning/refrigeration systems," *Applied Energy*, 2016.

[7] Z. Li, Z. Sun, Q. Meng, Y. Wang, and Y. Li, "Reinforcement learning of room temperature set-point of thermal storage air-conditioning system with demand response," *Energy and Buildings*, 2022.

[8] M. Mossolly, K. Ghali, and N. Ghaddar, "Optimal control strategy for a multi-zone air conditioning system using a genetic algorithm," *Energy*, 2008.

[9] R. Tang, S. Wang, K. Shan, and H. Cheung, "Optimal control strategy of central air-conditioning systems of buildings at morning start period for enhanced energy efficiency and peak demand limiting," *Energy*, 2018.

[10] C. Xiong, Z. Sun, Q. Meng, Z. Li, Y. Wei, F. Zhao, and L. Jiang, "A simplified improved transactive control of air-conditioning demand response for determining room set-point temperature: Experimental studies," *Applied Energy*, 2022.

[11] H. Dyvia and C. Arif, "Analysis of thermal comfort with predicted mean vote (pmv) index using artificial neural network," in *IOP Conference Series: Earth and Environmental Science*, vol. 622, no. 1. IOP Publishing, 2021, p. 012019.

[12] V. F. Ličina, T. Cheung, H. Zhang, R. De Dear, T. Parkinson, E. Arens, C. Chun, S. Schiavon, M. Luo, G. Brager *et al.*, "Development of the ashrae global thermal comfort database ii," *Building and Environment*, vol. 142, pp. 502–512, 2018.

[13] J. P. Holman, *Heat transfer*. McGraw Hill, 1986.

[14] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, "Human-level control through deep reinforcement learning," *nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[15] J. Shen, J. Cao, and X. Liu, "Bag: Behavior-aware group detection in crowded urban spaces using wifi probes," *The World Wide Web Conference*, 2019.

[16] X. Yu, Q. Zhang, J. Kang, and F. Cui, "Predicting integrated thermal and acoustic performance in naturally ventilated high-rise buildings using cfd and fem simulation," in *Building Simulation*, vol. 11. Springer, 2018, pp. 507–518.

[17] J. Kim, D. Song, S. Kim, S. Park, Y. Choi, and H. Lim, "Energy-saving potential of extending temperature set-points in a vrf air-conditioned building," *Energies*, vol. 13, no. 9, p. 2160, 2020.

[18] "Energy efficiency program for certain commercial and industrial equipment," 10 C.F.R. § 431.97, 2023.

Appendix A Further Details on our Simulation Setup

As mentioned above, we use the functions $Population(t)$, $In(t)$ and $Out(t)$ to model arrivals and departures. Fig. 3a and b provide a visualization, at specific times 11am and 3pm. The green square indicates the number of users who have entered the building, as well as their relative positions. The red square indicates the total population, from which users move and out into the building (the green square).

(a) 11:00 a.m.

(b) 3:00 p.m.

Figure 3: User Arrivals Simulation

The parameter settings for the environment and reward function calculation are as follows: The *EER* of variable refrigerant flow (VRF) air conditioning systems in large buildings is around 11 (USA CFR minimum energy performance standard, Table 15) [18]. The upper limit for the total daily mall population (6000), and the mall floor areas is derived according to [15]. The area of the space is set as $4890m^2$, and its height is set as $3 \times 4m$. The density of dry air is $1.275kg/L$. With this, we compute the mass as volume $\times$ density. The heat capacity of dry air is $1.00J/(gK)$.

We will release the simulator for other researchers, industry partners, and the public to use. The different components of the simulator can be modified, and customized for the specific building/public space setting. As an example, the simulation and environment can be modified to model real world scenarios like having more elderly people arrive in the morning and more young people arrive in the evening, where the elderly and young people may have different thermal comfort preferences.

Appendix B Further Details on Thermal Comfort Predictions

Modelling 9 thermal modes of users: We use the ASHRAE user thermal perception dataset [12]. There are 6 features which can be used to calculate PMV values: the metabolic rate (met), clothing insulation (clo), dry bulb air temperature, mean radiant temperature, relative humidity, and relative air velocity. From the dataset, we focused on the Summer season and obtained the interquartile values of the features 'met' and 'clo' at 25-th, 50-th, and 75-th percentiles. This 3 by 3 combination enables us to model 9 modes of users. In Table 1, we show statistics regarding the value of the features 'Metabolic rate' and 'Clothing insulation' used to calcuate the PMV.

Table 1: Characteristics of metabolic rate ('met') and clothing insulation ('clo') during **summer** in ASHRAE global thermal comfort dataset

	Metabolic rate	Clothing insulation
Count	37347	38010
Mean	1.225	0.575
Std	0.265	0.187
Min	0.65	0.00
25%	1.10	0.46
50%	1.20	0.57
75%	1.30	0.68
Max	6.83	2.55

In Fig. 4, we show how the PMV changes with temperature, for the 9 modes of users.

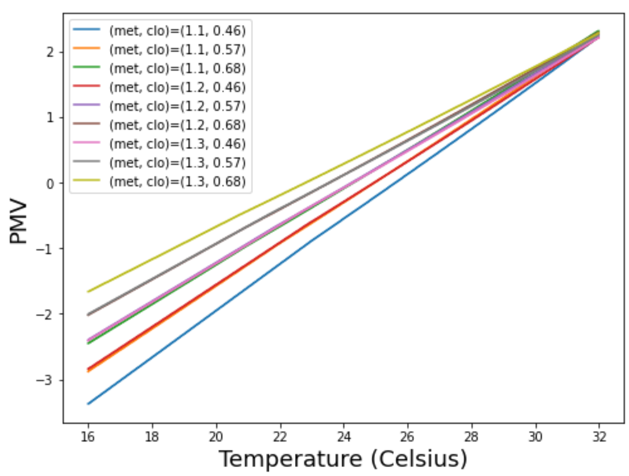


Figure 4: PMV as temperature changes, for 9 modes of users.

Appendix C Deep Q-Learning Algorithm Hyperparameter Setting

We set the following parameters for the RL algorithm within the HLAC RL framework.

Table 2: Hyper-parameters for HLAC RL

Name	Description
Batch size	128
Discount rate, γ	0.99
Initial ϵ	0.9
Final ϵ	0.05
ϵ decay	2000
Exploration rate decrease equation	$(0.9 - 0.05)e^{-x/2000}$
Update rate of the target network, τ	0.005
Learning Rate	0.0001