TUNEOPT: An Evolutionary Reinforcement Learning HVAC Controller For Energy-Comfort Optimization Tuning

Mostafa Meimand
Virginia Tech
Blacksburg, VA, USA
mostafaem@vt.edu

Vanshaj Khattar
Virginia Tech
Blacksburg, VA, USA
vanshajk@vt.edu

Zahra Yazdani
Georgia Tech
Atlanta, GA, USA
zahray@gatech.edu

Farrokh Jazizadeh
Virginia Tech
Blacksburg, VA, USA
jazizade@vt.edu

Ming Jin
Virgnia Tech
Blacksburg, VA, USA
jinming@vt.edu

ABSTRACT

HVAC systems account for the majority of energy consumption in buildings. Efficient control of HVAC systems can reduce energy consumption and enhance occupants' comfort. In the existing literature, energy-comfort or cost-comfort co-optimization frameworks commonly involve manual tuning of the balancing coefficient between energy and comfort through parameter tuning by an expert. Nevertheless, achieving the optimal balance between energy usage and occupant comfort remains challenging. This limitation restricts the generalizability of different formulations across various scenarios or testing on different environments. In this paper, we propose an implicit evolutionary Reinforcement Learning (RL) approach to learn and adapt the trade-off parameter of an energy-comfort optimization formulation. We have developed a predictive comfort-energy co-optimization formulation for controlling the setpoint of a building. The RL agent utilizes a novel guidance-induced random search method to learn the energy-comfort trade-off coefficient and guide the optimization formulation. The reward function of the RL model is energy productivity (comfort over energy consumption). To evaluate the feasibility of our proposed approach, we conducted experiments on a real-world testbed - i.e., an apartment unit. Our feasibility study shows that the proposed approach can learn an optimal control parameter and reduce energy consumption by 24.3% while decreasing comfort by only 1% compared to the baseline.

CCS CONCEPTS

• **Theory of computation → Sequential decision making**; **Random search heuristics**.

KEYWORDS

Reinforcement learning, Thermal comfort, Personal models, Energy-comfort trade-off, HVAC, Adaptive optimization, Parameter Tuning

BuildSys '23, November 15–16, 2023, Istanbul, Turkey

ACM ISBN 979-8-4007-0230-3/23/11.
https://doi.org/10.1145/3600100.3623751

ACM Reference Format:
Mostafa Meimand, Vanshaj Khattar, Zahra Yazdani, Farrokh Jazizadeh, and Ming Jin. 2023. TUNEOPT: An Evolutionary Reinforcement Learning HVAC Controller For Energy-Comfort Optimization Tuning . In *The 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '23), November 15–16, 2023, Istanbul, Turkey*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3600100.3623751

1 INTRODUCTION

Heating, Ventilation and Air Conditioning (HVAC) systems account for the majority of the energy use in buildings. Their control strategies seek to maintain occupants' comfort while accounting for efficient energy use. Optimization techniques are commonly studied and used for efficient control of HVAC systems [1, 2]. These approaches can encode domain-specific constraints and can handle problems with several decision variables [3]. Although these methods are well established with profound theoretical foundations, optimization formulations, once built, typically do not adapt to changing real-world conditions, such as occupants' differences and seasonal variations. This rigidity limits the flexibility of optimization approaches. As a common approach, optimization formulations are employed to minimize energy while balancing the trade-off between an energy use index and a measure of occupants' comfort. Typically, a trade-off coefficient between objective function terms is manually set through parameter tuning and used during operation. For instance, Kim [4] developed a model predictive controller to operate HVAC systems by considering individual thermal preferences. The model used a constant, C_{TD}, to balance energy cost and thermal discomfort. It was shown that different C_{TD} values could affect the energy-comfort trade-off that could be leveraged for demand response. Research has demonstrated that the trade-off coefficient directly impacts controller performance by favoring either energy or comfort. However, the effective strategies for configuring such weight coefficients remain to be determined. Typically, they are tuned by experts to ensure the resulting controller achieves high energy efficiency with limited impact on occupants' comfort. On the other hand, Reinforcement Learning (RL)–based approaches have been promising due to their ability to handle uncertainty and to continuously adapt to changing conditions. Examples of RL applications in HVAC control can be found in [5, 6]. The main themes of RL-based controllers for HVAC systems in the literature

center around RL models for shaving energy peak, utilizing passive thermal storage of buildings, and learning building thermodynamic behavior through interactions with environments [7, 8].

Most common approaches in RL often use Q-learning or actor-critic-based methods to learn the optimal policy [7]. Unlike previous RL frameworks, in this paper, we propose a novel evolutionary search (ES) algorithm with a guidance function based on state-action-trajectory data, which only accesses the environment through interactive samples (reward, states, etc.). The proposed approach combines the synergistic strength of optimization-based and RL-based approaches to adaptively learn the parameters of an optimization model using an RL agent, referred to as TUNEOPT (TUNE-OPTimization). Hence, instead of using the RL agent directly for taking actions, we use it to learn the parameters of an optimization model, while the control actions are taken by the optimization model. At its core, TUNEOPT leverages a predictive optimization formulation with the objective of minimizing energy consumption and maximizing occupant comfort while considering the HVAC system constraints. The RL agent guides the optimization formulation to maximize an energy productivity measure (comfort over energy). The proposed approach has been tested on a real-world apartment unit, and the results are compared against a baseline controller.

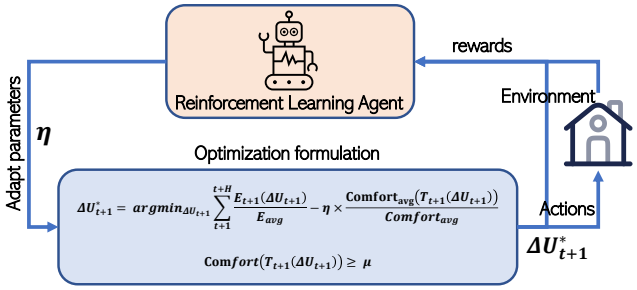

Figure 1: Basic architecture for TUNEOPT

2 METHODOLOGY

Figure 1 shows the TUNEOPT framework. In this framework, the RL agent adapts the η value to optimize operations and maximize the reward function, which is set as comfort over energy. In the optimization formulation, the parameter (η) serves as the balancing factor, dictating the trade-off between energy consumption and user comfort. The optimization formulation takes actions denoted as ΔU^*_{t+1}, which represents the change of setpoint, and communicates these actions to the environment through a thermostat. In the following subsections, we delve into the details of the optimization formulation and the RL agent.

2.1 HVAC Controller

A model predictive controller has been developed to co-optimize energy and comfort for a single-zone apartment, as demonstrated in Equation 1 and Equation 2. The first term is energy consumption (E_{t+1}) which is a function of temperature and the changing setpoint (ΔU_{t+1}) for the next H time steps. Energy consumption is estimated using a multivariate regression model explained in section 2.1.1. The second term pertains to thermal comfort, which is a function of indoor temperature (T_{t+1}). We have employed probabilistic personalized comfort models to accommodate individual thermal preferences, as these models offer accuracy at the individual level. Figure 2 illustrates an example of a comfort profile used in this study representing single occupancy. The corresponding occupant experiences 100% comfort around 73.9°F (23.3°C) and 50% comfort within the range of 70.6°F (21.4°C) to 77.1°F (25.1°C). More details of personalized comfort modeling could be found in [9]. Both terms are normalized using E_{avg} and $Comfort_{avg}$. The optimization is solved as a linear programming problem after linearizing the objective function. The parameter η governs the priority of comfort over energy, while μ in Equation 2 denotes the minimum probability of comfort that the controller must maintain. Although this formulation is for a single-zone environment, it can be expanded to multi-zone environments by changing the scalars to vectors.

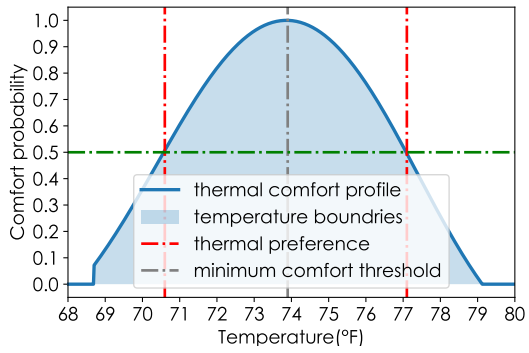

Figure 2: The corresponding thermal comfort profile

$$\Delta U^*_{t+1} = argmin_{\Delta U_{t+1}} \sum_{t+1}^{t+H} \frac{E_{t+1}}{E_{\text{avg}}} - \eta \cdot \frac{Comfort(T_{t+1})}{Comfort_{\text{avg}}} \quad (1)$$

$$s.t. \quad Comfort(T_{t+1}) \geq \mu \quad (2)$$

2.1.1 Building model. To model the building's thermodynamic behavior, a dataset representing thermal behavior variables including indoor and outdoor temperatures and HVAC setpoints is used. Two multivariate linear regression models are used for forecasting temperature (Equation 3) and energy consumption (Equation 5). In this formulation, T_t and ΔT_{t+1} represent the zone temperature at time t and the temperature change from time t to time $t+1$, respectively. Similarly, U_t and ΔU_{t+1} denote the setpoint and the change of setpoint from time t to time $t+1$. The $Dist$ vector comprises disturbances, including outdoor temperature and occupancy flags. The scalars a, b, and c are calculated through multi-variate regression analysis.

$$\Delta T_{t+1} = a_1 * T_t + b_1 * \Delta U_{t+1} + c_1 * Dist \quad (3)$$

$$\Delta T_{t+1} = T_{t+1} - T_t \quad (4)$$

$$E_{t+1} = a_2 * T_t + b_2 * \Delta U_{t+1} + c_2 * Dist \quad (5)$$

$$\Delta U_{t+1} = U_{t+1} - U_t \quad (6)$$

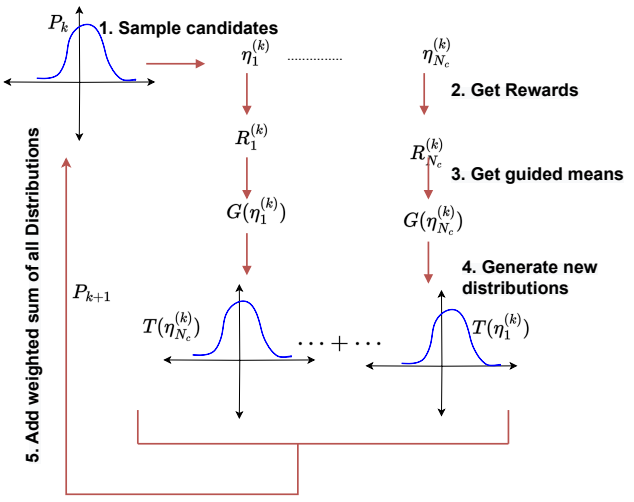

Figure 3: Visualization of the Algorithm 1.

2.2 Reinforcement Learning agent

The RL agent is designed to tune η for optimizing energy-comfort trade-offs. The reward function for the RL agent training is energy productivity, which is shown in Equation 7. The reward function quantifies the comfort achieved by consuming a unit of energy. To this end, the RL agent makes the sequential decisions for the parameter value η every day, such that the reward is maximized.

$$Energy\ productivity\ (r_t) = \frac{Comfort(\%)}{Energy\ consumption(kWh)} \tag{7}$$

2.3 Guided evolutionary search for parameters

The proposed evolutionary RL algorithm is shown in Algorithm 1. Firstly, we start with some N_c candidate parameter values for η, which are sampled from a probability distribution P_k. The probability distribution P_k can be determined from expert knowledge, e.g., a normal distribution for positive parameters, with mean and variance based on historical parameter values. We denote the index of parameter candidates with j and the iteration index with k. For each parameter candidate, we evaluate it on the environment and observe a noisy reward r_t. Then we use a predefined guidance function $G(.)$ for each candidate parameter. The guidance function $G(.)$ is to generate new distributions for each of the parameters, such that the mean of each distribution moves towards the best parameters observed in the current iteration k. In this work, we use the guidance function, which takes the mean of some of the best parameters observed in the current iteration. We guide the search for the parameters with this guidance function, which is a function of guidance factor ρ, and obtain new distributions for each candidate. Then a weighted sum of these distributions is taken on how the parameters are performed on the environment to get a new distribution P_{k+1} for the next iteration. See Figure 3 for the visualization of the algorithm. The algorithm stops when the improvement of the reward function falls below a certain threshold.

Algorithm 1: TUNEOPT Algorithm

1 N_c, Σ, ρ, $P_1 = \mathcal{N}_d(0, \Sigma^2)$, $k = 1$

2 k=1,2, $\cdots$ $Sample$ N_c candidates from the distribution P_k: $\eta_1^{(k)}, \eta_2^{(k)}, \cdots, \eta_{N_c}^{(k)}$ $j = 1, \ldots N_c$ Deploy the optimization actions for η_j in the environment. Observe reward $r_t(\eta_j^{(k)})$ from the environment

3 Sort the N_c candidates on the basis of rewards obtained. Let $\eta_1^\star$ and $\eta_2^\star$ be the best parameters. Calculate

$$A_g^{(k)} = mean(\eta_1^\star, \eta_2^\star)$$

4 Compute the guidance for each of the candidate by:

5

$$G(\eta_j^{(k)}, \rho) = \eta_j^{(k)} + \rho||A_g^{(k)} - \eta_j^{(k)}||_2. \tag{8}$$

Compute the new probability distribution P_{k+1} for the next:

6

$$P_{k+1} = \sum_{j=1}^{N_c} r_j^{(k)} T(\eta_j^{(k)}). \tag{9}$$

where:

$$r_j^{(k)} = \frac{\exp[r_t(\eta_j^{(k)})]}{\sum_{j=1}^{N_c} \exp[r_t(\eta_j^{(k)})]}. \tag{10}$$

7

$$T(\eta_j^{(k)}) = \mathcal{N}_d(G(\eta_j^{(k)}, \rho), \Sigma^2). \tag{11}$$

8 Increment $k \leftarrow k + 1$

3 REAL-WORLD TESTBED

Our real-world testbed is a one-bedroom apartment (655 SF) located in Blacksburg, VA. The air conditioning (AC) unit is controlled using an Ecobee smart thermostat. The control commands are sent to the thermostat via the Ecobee API [10]. Additionally, we monitored the energy consumption of the AC system using an Emporia Smart Home Energy Monitor [11], with a sampling rate of up to 1Hz. We used a 20-minute timestep by averaging the energy consumption data. The weather data were gathered from [12]. To generate the datasets for the predictive models, we randomly changed the setpoint between 70°F (21.1°C) to 77°F (25°C) during a three-day period. The Mean Absolute Error (MAE) for the trained models were 0.33°F and 0.037 kW for the temperature model (Equation 3) and the energy model (Equation 5), respectively. To acquire future outdoor temperature, we utilized the Meteomatics Python library [13]. The thermal comfort profile used was synthetically generated utilizing real-world data [9], where 100% comfort was at approximately 73.9°F as shown in Figure 2. The parameters employed for the optimization formulation included $H = 1$ and $\mu = 0.5$. The RL parameters were set to be $\rho = 1$, $N_c = 3$, and $\Sigma = \frac{4}{iteration\ number}$ as standard deviation. We compared the performance of TUNEOPT with a common engineering practice of manual tuning [14]. To this end, to choose the baseline, we established an initial distribution for η using expert knowledge. Subsequently, we randomly selected three η values and ran the optimization in the testbed on separate

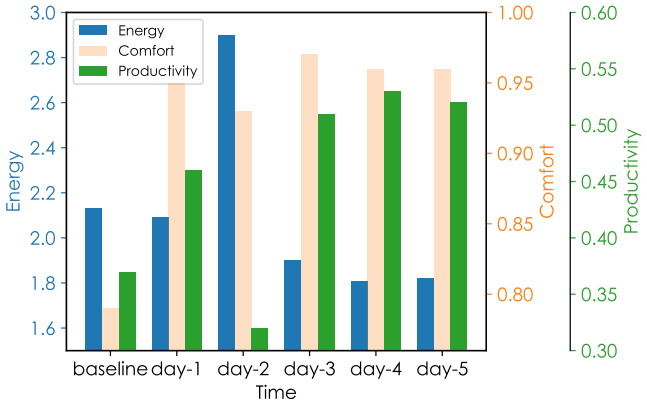

Figure 4: Evaluation metrics during experiment

days. Then, we determined the baseline η value based on the highest energy productivity achieved.

4 RESULTS AND DISCUSSION

Figure 4 illustrates energy consumption, comfort, and productivity throughout the experimental period. Note that the average outdoor temperature was between 72°F and 74°F during the experiment and baseline selection, which we assume has a negligible effect on the results. The algorithm begins with an initial Gaussian distribution, as proposed by an expert (step 1 in Algorithm 1). Next, three-parameter candidates for the balancing coefficient (η) are randomly sampled from this initial Gaussian distribution, and the controller is run for three days for each of these candidates (step 2). In step 3, the parameter candidates are sorted based on their corresponding reward values. The mean of the top two candidate parameters is used as input to the designed guidance function $G(\cdot)$. Based on the guidance function, a new distribution ($\mathcal{N}_d(G(\eta_j^{(1)}, \rho))$ is computed, and a new parameter candidate ($\eta_j^{(1)}$) is sampled for the next day (day-1) using the new distribution. At the end of day 1, the algorithm uses the reward value from day 1 along with the reward values from baseline to estimate a new distribution ($\mathcal{N}_d(G(\eta_j^{(2)}, \rho))$ and samples a new parameter value ($\eta_j^{(2)}$). The training process continues until day 5 when the rewards are no longer improving. This study shows TUNEOPT's feasibility for adaptive energy optimization. Future research should extend the experiment period, assess the baseline controller's performance across multiple days and occupancy scenarios, and test the proposed controller with various baseline controller seeds. In light of the limitations, the summary outcome of the feasibility study is shown in Table 1. The TUNEOPT controller achieves a 32.5% improvement in energy productivity and a 24.3% reduction in energy consumption, with only a marginal 1% compromise in comfort.

5 CONCLUSION

This paper presents TUNEOPT, an evolutionary reinforcement learning (RL) HVAC system controller designed to adapt and fine-tune an energy-comfort co-optimization controller in a dynamic environment, enabling dynamic responses to changing real-world conditions. The RL agent seeks to maximize a reward function by tuning the predictive controller. Through real-world testbed experiments on an apartment unit, the feasibility of the TUNEOPT was demonstrated in learning and improving energy efficiency (by reducing 24% of energy use). As a future research direction, the extension of experiments, assessment of the baseline controller's performance over multiple days, exploration of TUNEOPT's adaptability to diverse climates and seasons, and evaluation of its performance in complex multi-occupancy scenarios with multi-variable tuning could be pursued.

Table 1: Results: TUNREOPT vs. baseline controller

	TUNEOPT	baseline	Change(%)
Comfort (%)	0.95	0.96	1% ↓
Energy (kWh)	1.80	2.38	24.3% ↓
Energy Productivity	0.53	0.4	32.5% ↑

6 ACKNOWLEDGMENTS

MM and FJ acknowledge the support from Virginia Tech's BioBuild program. VK and MJ acknowledge the support from NSF grant #2034137, C3.ai Digital Transformation Institute, and the U.S. Department of Energy. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the listed funding agencies.

REFERENCES

[1] Ali Ghahramani, Farrokh Jazizadeh, and Burcin Becerik-Gerber. A knowledge based approach for selecting energy-aware and comfort-driven hvac temperature set points. *Energy and Buildings*, 85:536–548, 2014.
[2] Mostafa Meimand and Farrokh Jazizadeh. Human-in-the-loop model predictive operation for energy efficient hvac systems. In *Construction Research Congress 2022*, pages 178–187, 2022.
[3] Francisco Facchinei and Jong-Shi Pang. *Finite-dimensional variational inequalities and complementarity problems*. Springer, 2003.
[4] Young-Jin Kim. Optimal price based demand response of hvac systems in multi-zone office buildings considering thermal preferences of individual occupants buildings. *IEEE Transactions on Industrial Informatics*, 14(11):5060–5073, 2018.
[5] Guanyu Gao, Jie Li, and Yonggang Wen. Deepcomfort: Energy-efficient thermal comfort control in buildings via reinforcement learning. *IEEE Internet of Things Journal*, 7(9):8472–8484, 2020.
[6] Bingqing Chen, Zicheng Cai, and Mario Bergés. Gnu-rl: A precocial reinforcement learning solution for building hvac control using a differentiable mpc policy. In *Proceedings of the 6th ACM international conference on systems for energy-efficient buildings, cities, and transportation*, pages 316–325, 2019.
[7] Zoltan Nagy, Gregor Henze, Sourav Dey, Javier Arroyo, Lieve Helsen, Xiangyu Zhang, Bingqing Chen, Kadir Amasyali, Kuldeep Kurte, Ahmed Zamzam, et al. Ten questions concerning reinforcement learning for building energy management. *Building and Environment*, page 110435, 2023.
[8] Vanshaj Khattar and Ming Jin. Winning the citylearn challenge: Adaptive optimization with evolutionary search under trajectory-based guidance. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 14286–14294, 2023.
[9] Wooyoung Jung and Farrokh Jazizadeh. Comparative assessment of hvac control strategies using personal thermal comfort and sensitivity models. *Building and Environment*, 158:104–119, 2019.
[10] ecobee API — ecobee.com. https://www.ecobee.com/home/developer/api.
[11] Emporia: Revolutionizing Home Energy — emporiaenergy.com.
[12] Blacksburg, VA Past Weather For Last 30 days - LocalConditions.com — local-conditions.com. https://www.localconditions.com/weather-blacksburg-virginia/24060/past.php. [Accessed 24-Jul-2023].
[13] Python | Meteomatics — meteomatics.com. https://www.meteomatics.com/en/api/data-connectors/python/.
[14] Marco Biemann, Fabian Scheller, Xiufeng Liu, and Lizhen Huang. Experimental evaluation of model-free reinforcement learning algorithms for continuous hvac control. *Applied Energy*, 298:117164, 2021.