Optimal Room Temperature Control using Reinforcement Learning

Aigerim Gilmanova
Robotics and Computer Vision
Innopolis Unversity
Innopolis, Republic of Tatarstan
a.gilmanova@innopolis.university

Ahror Jabborov
Data Analysis and Artificial Intelligence
Innopolis Unversity
Innopolis, Republic of Tatarstan
a.jabborov@innopolis.university

***Abstract*—In developed countries, almost half of the energy consumption is account by the climate control systems such as heating, ventilation and air conditioning. These systems, called as HVAC, require a lot of electricity and are the main causes of increased demand. Traditional ways for optimizing the energy usage by these systems usually result in decreased amount of human satisfaction because people prefer more desirable and comfortable conditions. Therefore, such situations create a significant demand for the intelligent optimization approaches. Intelligent scheduling of operational times of each HVAC system can help achieve the desired significant reduce in the energy costs. However, complexity in building's thermal dynamics and environment disturbances has the power to decrease the efficiency. One great solution is to transfer the favorable conditions from one side of the building to the other parts so that we can avoid the energy wastage in different parts of buildings. In this paper, we study a simple scenario of heat transferring between rooms. We develop a suitable simulator with corresponding environment to be able to train our agent. We evaluate our agent based on our reward function which considers the favorable conditions inside rooms.**

***Index Terms*—HVAC (Heating, ventilation, air conditioning) system, temperature control, reinforcement learning, q-learning.**

I. Introduction

Nearly 40% of the total energy is consumed by simple temperature control systems in buildings and households [1]. Intelligent thermal control inside the building allows it to shift the energy demand of HVAC systems while maintaining the desired levels of room temperatures [2]. Unlike developing countries, most of the developed countries charge consumers bills based of time-of-use price which means that the prices of energy changes frequently based on the demand level at a corresponding time period. Therefore, thermal conditions control provides great potential solution for decreasing the energy bills and improving the efficiency and stability levels of the grid avoiding the likelihood of energy outages and possible problematic scenarios. In this paper, we focus on the thermal shifting of temperature between rooms. For this purpose, we develop a simulation with 3 rooms and train our agent to optimize the temperature control to achieve desired level. We use Q-learning because it has the potential to converge towards the optimal policy by finding the best course of action, given the current state of the agent.

II. Problem Formulation

Controlling the thermal conditions of rooms can help achieve significant energy usage optimizations. Heat transferring between rooms play an important role in achieving the temperature control in buildings. Therefore, the main focus of the following project is controlling the room temperatures by shifting the heat from one room to another. It allows to save energy costs by focusing on the maintaining the desired room temperature. For simulation cases, this work will put the room temperature in higher importance than the energy optimization.

III. Literature Review

This section summarizes the literature that was reviewed considering the HVAC system control by using reinforcement learning approaches. In the academia, based on the statistical data, this field has been studied thoroughly during the last 5 years when energy costs showed a significant rise [3], [4], [5], [6]. Majority of the researchers focused on approaches which use a simplified model for controlling the thermal dynamics during operational times in order to predict the building's temperature changes. Y. Ma and et al. [3] mainly focused on cooling systems such as cooling towers, storage banks and chillers and worked on a nonlinear model in order to develop a predictive control model scheme to optimize the energy consumption. Oldewurtel and et. el. [5] worked on a model to control the optimization of thermal conditions and they treat the problem as sequential linear programming (SLP). M. Maasoumy and et al. [4] mainly studied the dynamics of thermal control and developed tracking linear-quadratic regulator (LQR). Similarly, T. Wei, Q. Zhu, and M. Maasoumy [6] worked on building a model based on MPC-based algorithm in order to control the operating schedules of HVAC systems to cooperate with other demands and supplies. The authors of these papers claim to have significant results but the thermal dynamic model has to be efficiently addressed using mathematical tools for practical run-time control because it plays an important role on the performance and the reliability of the proposed approaches [7]. Nevertheless, there are a lot of factors which significantly affects the temperature levels of buildings, e.g. outside climate, building walls' materials

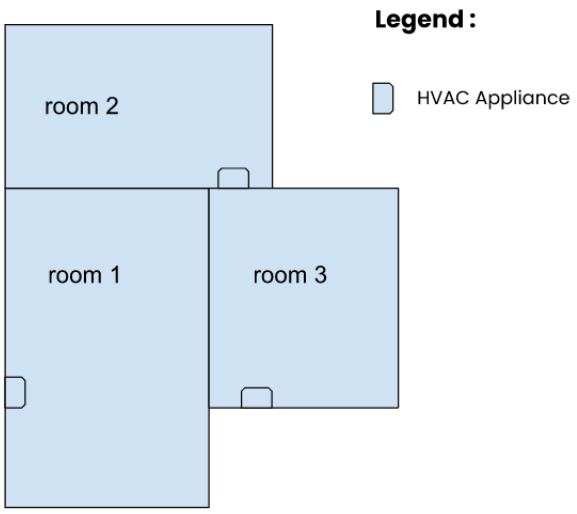
and internal heat consuming by occupants. Therefore, usually the thermal conditions inside buildings show random behaviours under incomplete modeling. For this reason, a lot of researchers have started to develop data-driven approaches to use real-time data for their RL methods. The authors of [8], [9] and [10] use classical Q-learning techniques which include tabular Q value function. The authors of [11] worked on optimal temperature setting point using neural fitted RL method. However, this approach is tested on single-zone buildings and they used simple forms of differential equations to model the heat transfer. The authors of [12] focused on developing a model-assisted batch RL method. They used randomized trees in order to approximate the action value and make appropriate turning on and off control strategy decisions. These approximation methods can work for state situations which are continuous while requiring extensive computing power for training and learning processes.

IV. Methodology

The following section describes the approach that was used in the implementation for the HVAC system optimal temperature control.

A. Environment

The Environment of the project is a simple model which includes a heat transfer between rooms and the exterior. There are three rooms which are connected in the configuration, which is shown below.



1) The model: The heat transfer is calculated as follows:

- T_1, T_2 and T_3 are the temperatures in room 1, 2 and 3, respectively.
- T_{ext} is the exterior temperature.
- $in_{transfer}$ is a factor that represents the speed of the heat transfer inside the building.
- $ext_{transfer}$ is a factor that shows the speed of the heat transfer between the building and the exterior.

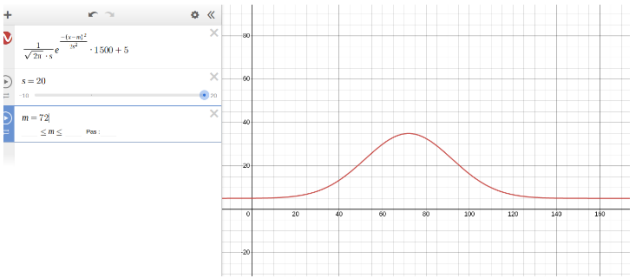
Every ten minutes, the temperature in each room will update as such :

- $T_1 = T_1 + in_{transfer} \times((T_2 + T_3)/2 - T_1) + ex_{transfer} \times(T_{ext} - T_1)$
- $T_2 = T_2 + in_{transfer} \times((T_1 + T_3)/2 - T_2) + ex_{transfer} \times(T_{ext} - T_2)$
- $T_3 = T_3 + in_{transfer} \times((T_1 + T_2)/2 - T_3) + ex_{transfer} \times(T_{ext} - T_3)$

The following model captures in essence the dynamics of temperature evolution on a building. A room temperature is updated using two factors:

- The difference of temperature between the room and the average temperature of adjacent rooms.
- The difference of temperature between the room and the exterior.

2) Exterior Temperature: The exterior temperature will change throughout the day between 5°C and 35°C, following the function below:



B. State Space

The state space of the agent is defined by the temperature in each room. In order to apply Tabular Q-learning, we need a finite number of states, we therefore round the actual temperature of the room to the closest integer. Furthermore, for the Q-learning algorithm to be efficient with our available time and computing resources, we should limit the possible number of states. We consider a frame of 5°C around the optimal temperature, 20°C.

- If the temperature given by the model in a given room is lower than 15°C, the composant of the state vector for this given room will stay at 15.
- If the temperature given by the model in a given room is higher than 25°C, the composant of the state vector for this given room will stay at 15.

This is not an issue, as if the agent's state reaches (15,15,15) or (25,25,25), the reward of the agent will be the worst achievable (see **Reward**). The agent should therefore quickly learn to stay far from those states.

Therefore, we have in our context $11^3 = 1331$ possible states. This is very reasonable.

C. Reward

We define the optimal temperature in a room to be 20°C. The reward is simply the negative distance between the optimal temperature vector (20,20,20) and the state of the agent. We choose the Euclidean distance for this problem. Therefore, the highest reward is 0, when the temperature is optimal and the lowest reward is minus square root of 75, around -8.66.

D. Action Space

The agent controls the HVAC system. It can therefore turn on or off the heating and cooling appliances in each room. We consider the following possible actions:

- Turn on the heating, represented by 1.
- Turn on the cooling, represented by -1.
- Turn off the system, represented by 0.

An action is therefore represented by a vector (X,Y,Z) where X,Y,Z can take the values (0, -1, 1). For 3 HVAC controllers, this leads to $3^3 = 27$ possible actions.

We consider that the actions are doing the following on the environment.

- If the heating is turned on, the room is heated by 1 C° after ten minutes.
- If the cooling is turned on, the room is cooled by 1 C° after ten minutes.
- If the appliance is turned off, the room temperature does not change actively, only the dynamic of the environment influences it.

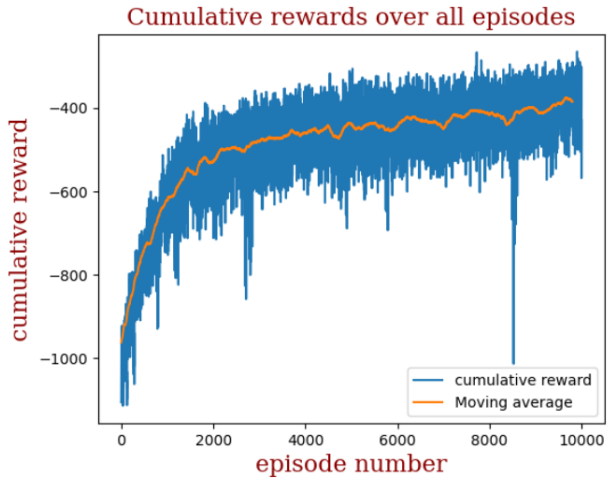
E. Episodic Context

We train our agent in an episodic setting, considering 24 hours time frame, divided into ten minute steps. Therefore, we consider episodes with a length of 144 steps.

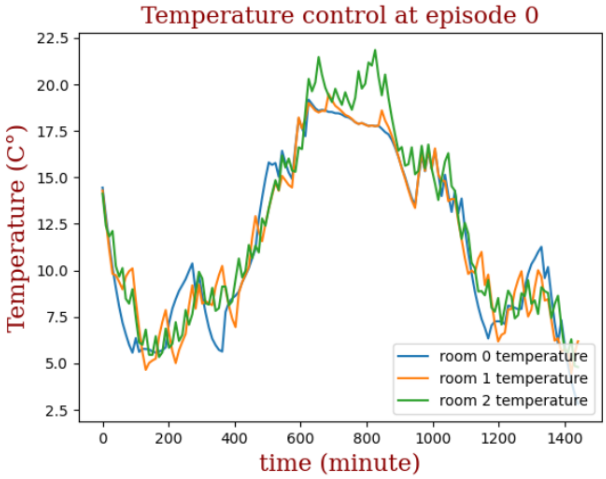
Computing the cumulative reward for each episode will allow us to evaluate the agent's ability to learn.

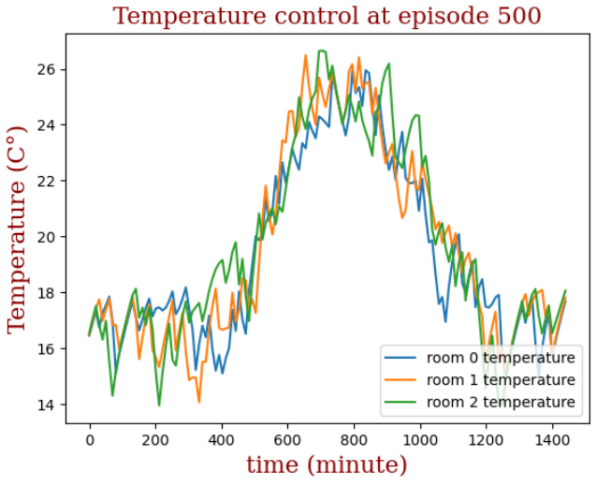
V. Results and Discussion

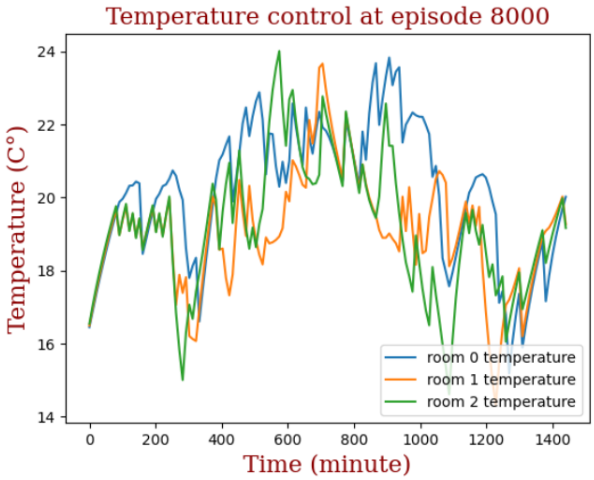
The plot below shows the result of a Q-Learning training during 10000 episodes, each composed of 144 steps. For this training, a constant exploration factor was chosen, epsilon = 0.2, as well as a constant discount factor = 0.95 and a constant learning rate of 0.1. We clearly see on the plot an effective learning, stabilizing after 4000 episodes. The moving average window is of size 200.



On the three plots below, we see how the temperature in each room changes in a 24 hours period, after training for 0 episodes, 500 episodes and 8000 episodes.







We can clearly see on the first plot that the HVAC system is not very effective, the temperature drops close to 5 degrees and reaches around 22.5 degrees. On the contrary, on the third plot, after 8000 episodes of training, the system is much more efficient. The temperature never drops under 14 degrees and never gets above 24 in the three rooms.

Tweaking the parameters of the Q Learning algorithm might improve further the efficiency of the learning. In particular, the

sum of the learning rates must diverge, but the sum of their squares must converge, for Q-Learning to converge toward the optimal policy.

VI. Conclusion

In conclusion, it can be said that the agent's ability to learn is effective, and it stabilizes after 4000 episodes. Thus, it is possible to control the room temperature in order to save to cost energy. Also, for the future work some further improvements can be considered.

- We could expand from 3 rooms to more, by using graphs to represent the connections between each room. The model will then take into account the temperature of the adjacent room to update itself.
- Using a better model for heat transfer between rooms, using methods such as the finite element method to solve a heat equation.
- Starting with a high learning rate and decreasing it over time might speed up the learning.
- Currently, the exploration factor epsilon is constant. Decreasing it slowly from a high value down to 0 over time, will allow exploration in the beginning and exploitation when the Q-Learning algorithm finds a near optimal policy.
- Q-Learning is a model free reinforcement learning algorithm, which is great when the environment's dynamics are too complex to model. In our case, good models of the evolution of temperature in a building exist, and we could incorporate such a model to make the learning more efficient. In this case, we would need to switch to model based RL algorithms.

References

[1] JL Blue, KH Lowe, BJ Hurlbut, et al. *Buildings energy use data book. Edition 2*. Tech. rep. Oak Ridge National Lab., TN (USA), 1979.

[2] Simon J Olivieri, Gregor P Henze, Chad D Corbin, et al. "Evaluation of commercial building demand response potential using optimal short-term curtailment of heating, ventilation, and air-conditioning loads". In: *Journal of Building Performance Simulation* 7.2 (2014), pp. 100–118.

[3] Yudong Ma, Francesco Borrelli, Brandon Hencey, et al. "Model predictive control for the operation of building cooling systems". In: *IEEE Transactions on control systems technology* 20.3 (2011), pp. 796–803.

[4] Mehdi Maasoumy, Alessandro Pinto, and Alberto Sangiovanni-Vincentelli. "Model-based hierarchical optimal control design for HVAC systems". In: *Dynamic Systems and Control Conference*. Vol. 54754. 2011, pp. 271–278.

[5] Frauke Oldewurtel, Alessandra Parisio, Colin N Jones, et al. "Energy efficient building climate control using stochastic model predictive control and weather predictions". In: *Proceedings of the 2010 American control conference*. IEEE. 2010, pp. 5100–5105.

[6] Tianshu Wei, Qi Zhu, and Mehdi Maasoumy. "Co-scheduling of HVAC control, EV charging and battery usage for building energy efficiency". In: *2014 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*. IEEE. 2014, pp. 191–196.

[7] Lei Yang, Zoltan Nagy, Philippe Goffin, et al. "Reinforcement learning for optimal control of low exergy buildings". In: *Applied Energy* 156 (2015), pp. 577–586.

[8] Enda Barrett and Stephen Linder. "Autonomous hvac control, a reinforcement learning approach". In: *Joint European conference on machine learning and knowledge discovery in databases*. Springer. 2015, pp. 3–19.

[9] Bocheng Li and Li Xia. "A multi-grid reinforcement learning method for energy conservation and comfort of HVAC in buildings". In: *2015 IEEE International Conference on Automation Science and Engineering (CASE)*. IEEE. 2015, pp. 444–449.

[10] D Nikovski, J Xu, and M Nonaka. "A method for computing optimal set-point schedules for HVAC systems". In: *Proceedings of the 11th REHVA World Congress CLIMA*. 2013.

[11] Pedro Fazenda, Kalyan Veeramachaneni, Pedro Lima, et al. "Using reinforcement learning to optimize occupant comfort and energy usage in HVAC systems". In: *Journal of Ambient Intelligence and Smart Environments* 6.6 (2014), pp. 675–690.

[12] Giuseppe Tommaso Costanzo, Sandro Iacovella, Frederik Ruelens, et al. "Experimental analysis of data-driven control for a building heating system". In: *Sustainable Energy, Grids and Networks* 6 (2016), pp. 81–90.