Optimal room temperature control using reinforcement learning

Aigerim Gilmanova
Ahror Jabborov

Outline

- Introduction on the area of your project
- Problem formulation
- Literature review
- Methodology
- Results and Discussion
- Conclusion

Introduction

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 to optimizing the energy usage by these systems usually results in decreased amount of human satisfaction because people prefer more desirable and comfortable conditions. Therefore, such situations create a significant demand for 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.

Problem Formulation

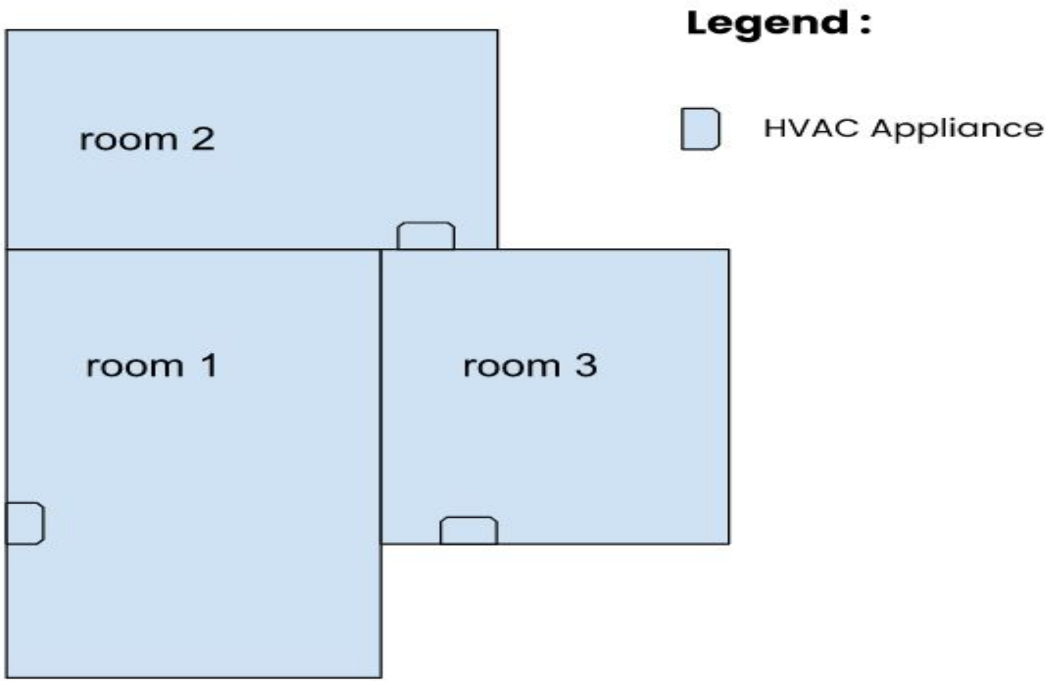
Controlling the thermal conditions of rooms can help achieve significant energy usage optimizations. Heat transferring between rooms play an important role in achieving the temperature control in buildings. Therefore, the main focus of the following project is controlling the room temperatures by shifting 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.

Literature Review

- 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.
- 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.
- 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.
- Q-learning by far is the most widely used based on the literature review.
- Some of the data-driven approaches use function approximation which can have high computational cost.
- Deep Reinforcement learning shows a great potential in terms of efficiency and performance.

Methodology

Environment



Methodology

The model

The heat transfer is computed as such :
Let's call T1, T2, T3 the temperature in room 1, 2, 3 respectively.
Let's call T_ext the exterior temperature.
Let's call in_transfer a factor representing the speed of the heat transfer inside the building.
Let's call ext_tranfer a factor representing the speed of the heat transfer between the building and the exterior.
Every ten minutes, the temperature in each room will update as such :

- T1 = T1 + in_transfer * ((T2+T3)/2 -T1) + ext_tranfer * (T_ext - T1)
- T2 = T2 + in_transfer * ((T1+T3)/2 -T2) + ext_tranfer * (T_ext - T2)
- T3 = T3 + in_transfer * ((T1+T2)/2 -T3) + ext_tranfer * (T_ext - T3)

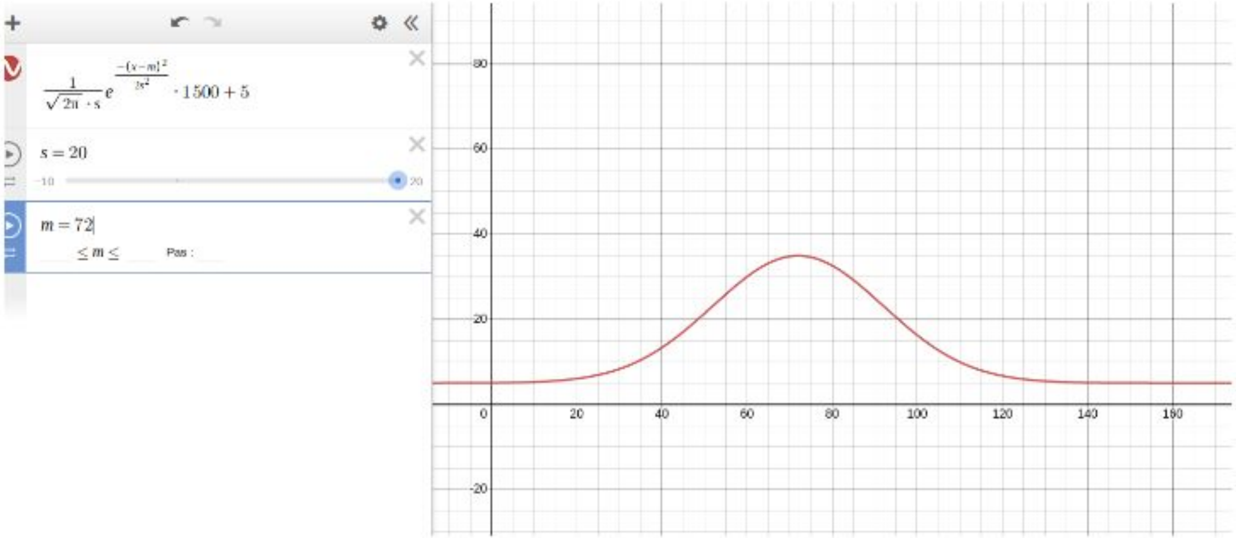
A room temperature is updated using two factors :

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

Methodology

Exterior temperature

The exterior temperature will change throughout the day between 5°C and 35°C, following the function below.



Methodology

State space

The state space of the agent is defined by the temperature in each room.
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 Q-learning 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.

We therefore have in our context 11^3 = 1331 possible states. This is very reasonable.

Methodology

Reward

- We define the optimal temperature in a room to be 20ºC.
- 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.
- The highest reward is therefore 0, when the temperature is optimal and the lowest reward is minus square root of 75, around -8.66.

Methodology

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.

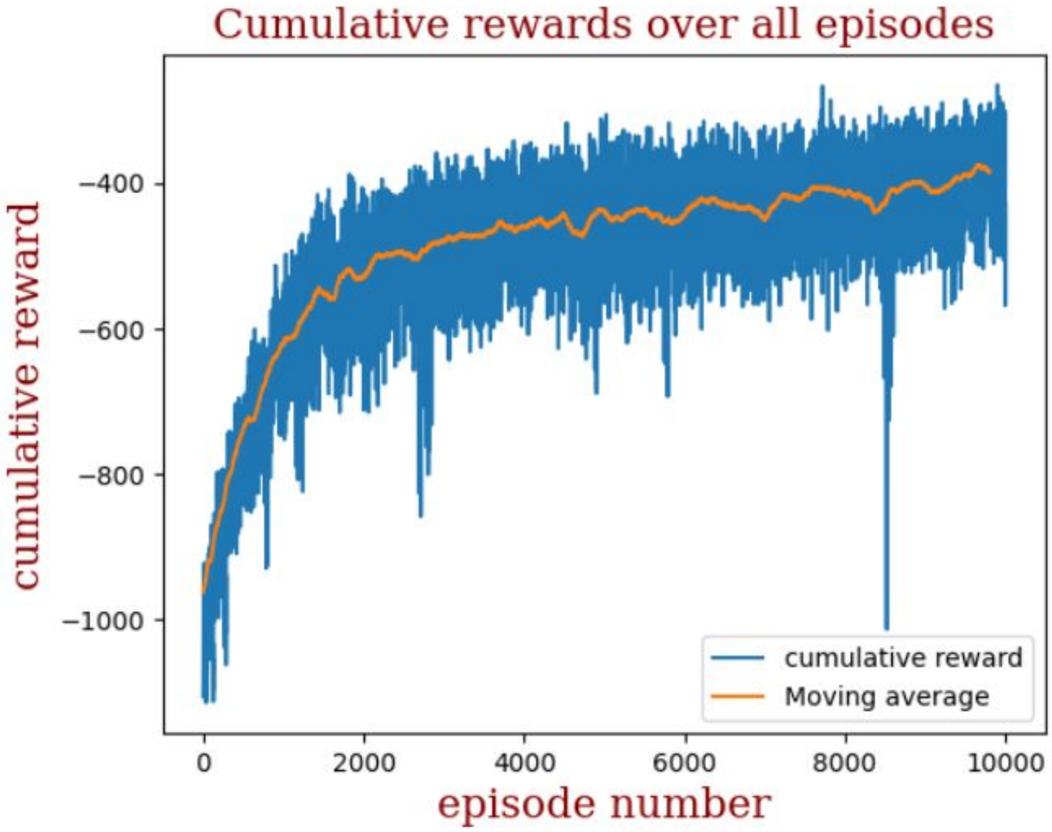
- If the heating is turned on, the room is heated by 1 C° after ten minutes.
- If the cooling is turned on, the room is cooled by 1 C° after ten minutes.
- If the appliance is turned off, the room temperature does not change actively, only the dynamic of the environment influences it.

Methodology

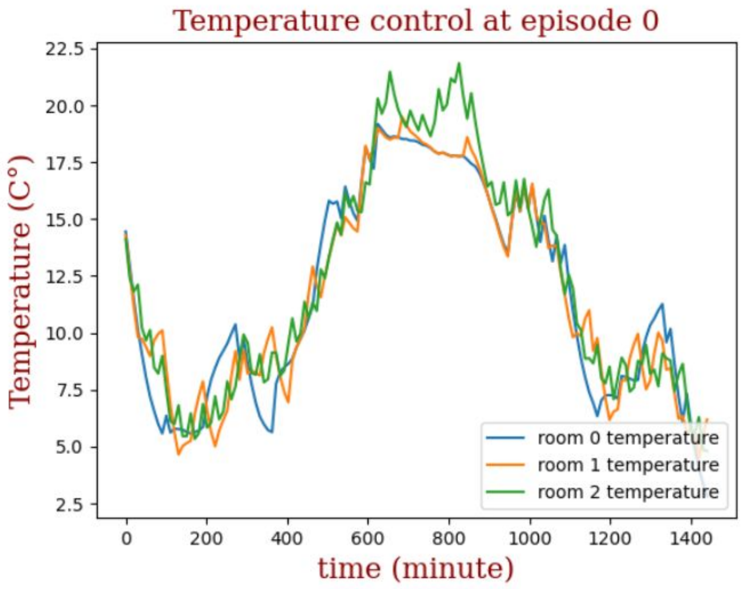
Episodic context

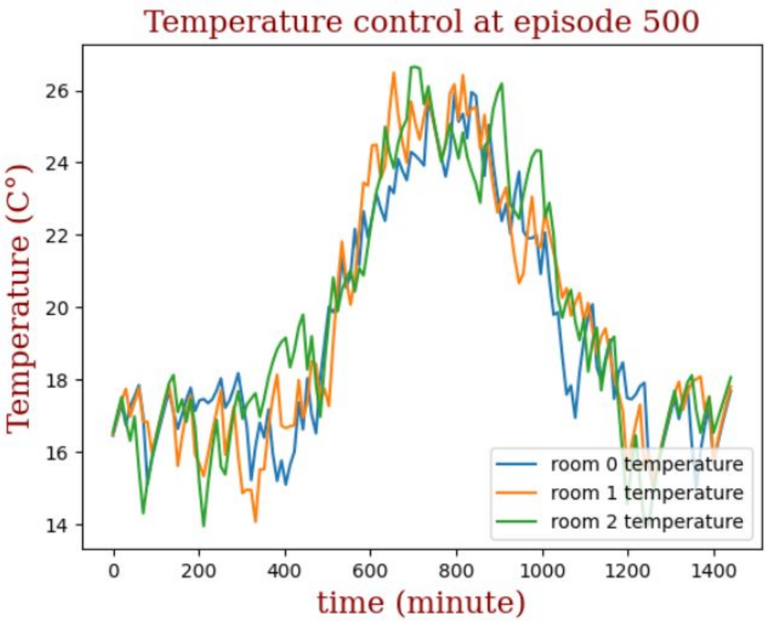
- We train our agent in an episodic setting, considering a 24 hours time frame, divided into ten minute steps. We therefore consider episodes with a length of 144 steps.
- Computing the cumulative reward for each episode will allow us to evaluate the agent's ability to learn.

Results and Discussion

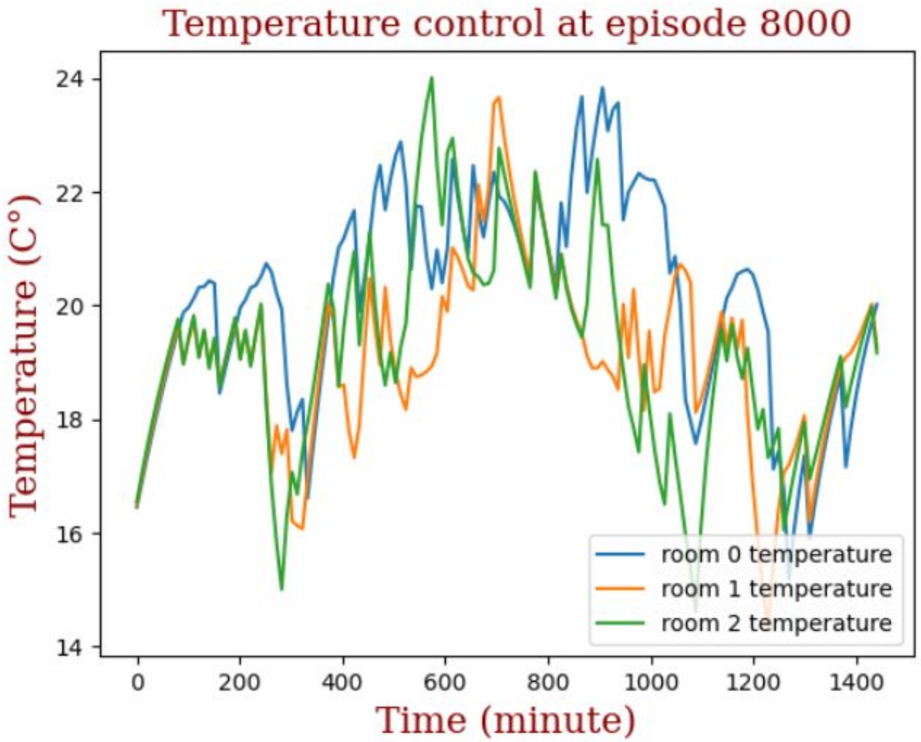


Results and Discussion





Results and Discussion



Conclusion

- 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 QLearning algorithm finds a near optimal policy.

- QLearning 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.

Contribution

Aigerim

- Literature review
- Model building
- Code implementation

Ahror

- Literature review
- Simulation
- Code implementation