ENAC / PROJET DE MASTER 2020-2021
SECTION DE GÉNIE CIVIL

Implementation and testing of a smart personalized thermostat with the help of a deep deterministic policy gradient algorithm

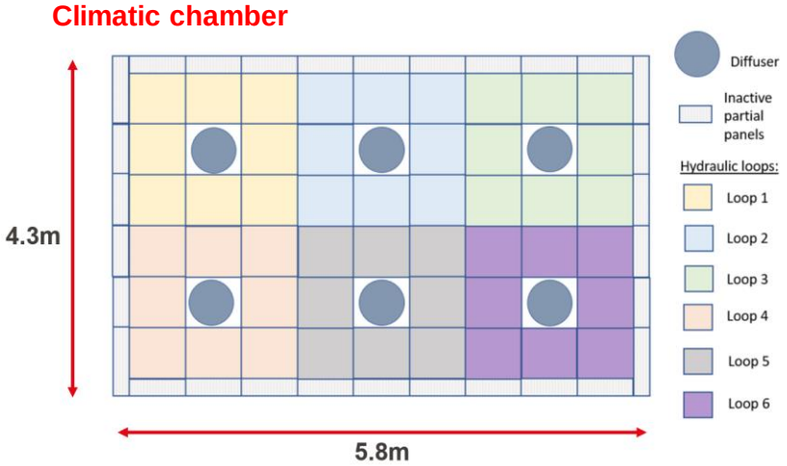
Auteur(e)s : Leonard Ung

Encadrement : Prof. Khovalyg Dolaana [1] / Ass. Doct. Chatterjee Arnab [1]

[1] Integrated Comfort Engineering (ICE)

Indoor air quality inside buildings is important for occupants to improve productivity and reduce health concerns. Especially in the last two years, due to the Corona virus and other diseases, one might think that if we increase the indoor air quality and indoor comfort, we will consume more energy. But with a proper indoor air control, it is even possible to decrease the energy consumption while increasing the thermal comfort. But nowadays, most of the thermostats maintain a static thermal environment which is not suitable for each person at each period of the day. The main purpose of our work is to develop a controller that is able to dynamically change the thermal environment based on occupants while optimizing the energy consumption.

Climatic chamber



- Air supply can come either from the ceiling or the floor by 6 diffusers
- 12 independent radiant heating or cooling panels (6 loops on the ceiling, 6 loops on the floor).

Control only with the radiant panels

- Open the ceiling valve at 100% if ΔT > 0.5
- Open the ceiling valve at 80% if 0.5 > ΔT > 0.1
- Open the ceiling valve at 60% if 0.1 > ΔT > 0.02
- Open the ceiling valve for 10min × ΔT with a minimum of 2min
- Keep the valve close for at least 6min
- With ΔT = (setpoint temperature – room temperature)

Deep Deterministic Policy Gradient (DDPG)

- RL is a part of machine learning and consist of an autonomous agent that learns the best action to take at a particular state, based on experience, by trying to maximize the reward at each iteration.
- DDPG is a deep reinforcement learning algorithm that is specifically adapted for environments with continuous action spaces (in our case the temperature).
- this algorithm learns from a Q-function and a policy
- The policy is then learned by the Q-function
- The Q-function is learned through off-policy data and the Bellman equation
- Actor-critic methods have two networks: actor network and critic network, which are shown in Fig 1

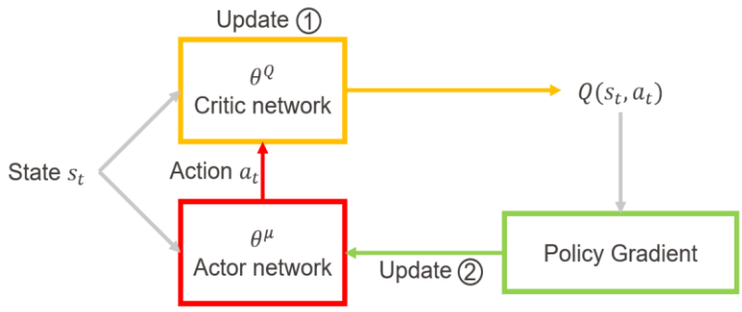
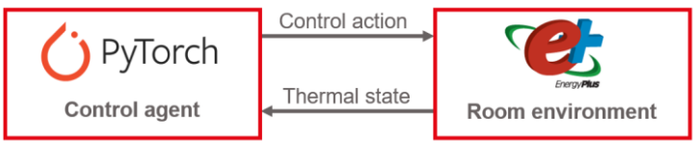


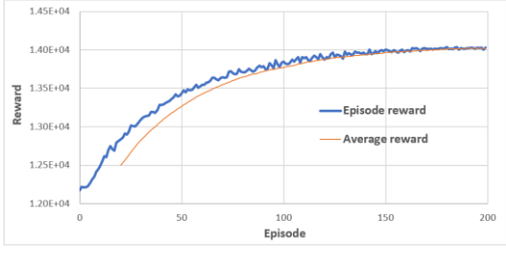
Fig 1 - DDPG flow chart

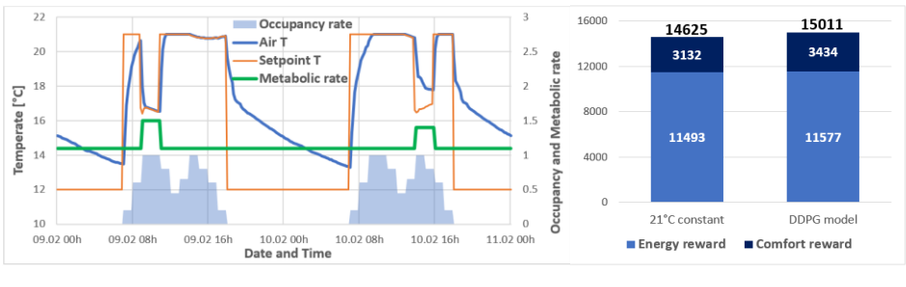


- Since we need a large amount of training data, the only way to perform the training in a reasonable time is to first train the DDPG algorithm on an *EnergyPlus* building model
- We perform a co-simulation between Python (*PyTorch*) and *EnegyPlus*

Results and performance

- During the 2 first episodes, only random actions were taken. We observe that the episode reward improves at each episode and finally stabilizes after around 150 episodes.
- We used the epsilon-greedy algorithm. Probability of ε to make random action and 1-eps of making the action it actually chose
- ε starts with a value of 1 and decay with a rate of 0.97 a each new episode. For the last episode, only 0.2% of the actions were taken randomly





- Variation of occupancy and metabolic rate (Met)
- Only heat when room is occupied
- Lower temperature when Met is rising
- Try to keep the PMV (Predicted Mean Vote) around -0.5
- 10% thermal comfort improvement compare to a standard thermostat
- Other algorithms (occupancy prediction) could increase reward