Deep Reinforcement Learning based HVAC Control for Thermal Comfort and Energy Efficiency in Office Buildings

Zhiang Zhang, Chenlu Zhang, Siliang Lu, School of Architecture

Introduction

Occupant thermal comfort and energy efficiency are the two main objectives for the operation of modern office buildings. However, conventional rule-based control strategies of the heating, ventilaion and air conditioning (HVAC) systems often lead to thermal discomfort and high energy consumption. This project aims at using deep reinforcement learning to control HVAC systems of a typical office building to achieve better thermal comfort and energy efficiency.

Training and Testing Environments

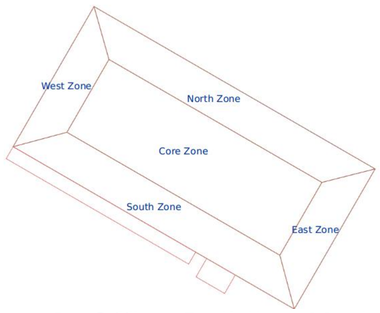

Figure 1: Office Building Model

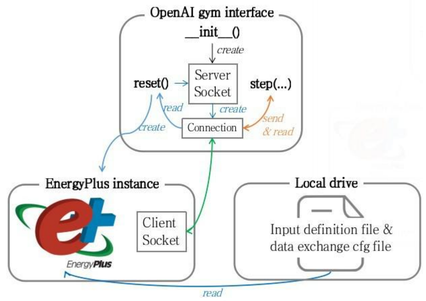

Figure 2: EnergyPlus and OpenAI gym

- EnergyPlus [1] is used as the building and HVAC systems simulator
- Typical single-level small office building model by US. DOE [2]
- Centralized variable air volume (VAV) system with terminal reheat
- **Training environment:** deterministic occupancy/equipment load schedule, Pittsburgh PA weather
- **Testing environment:** stochastic occupancy/equipment load schedule [3], State College PA weather, oversized HVAC equipment, slightly different centralized air handling unit control

State, Action and Reward Design

- **State:** weather conditions + environment conditions of the ***south zone*** + HVAC **total** power
- **Action:** adjustment to the ***south zone*** heating and cooling setpoint(C)
  - Default: *Zip*{(0.0, 1.0, -1.0, -1.0), (0.0, 1.0, -1.0, 1.0)}
  - Exp1: *Zip*{(0.0, 1.0, -1.0, -1.0, 2.0, -2.0, -2.0, 3.0, -3.0, -3.0), (0.0, 1.0, -1.0, 1.0, 2.0, -2,0, 2.0, 3.0, -3.0, 3.0)}
  - Exp2: *Zip*{(0.0, 1.0, -1.0, -1.0, 2.0, -2.0, -2.0, 5.0, -5.0, -5.0), (0.0, 1.0, -1.0, 1.0, 2.0, -2,0, 2.0, 5.0, -5.0, 5.0)}
- **Reward:** (PPD: predicted percentage of dissatisfied, thermal comfort metric, 0 means completely comfortable)
  - Linear: -λ * PPD - (1-λ) * HVACTotalPower, where λ∈[0, 1]
  - L2: -||PPD, HVACTotalPower||$_2$

Asynchronous Advantage Actor Critic (A3C) [4]

- **minibatch size:** 5, **state history stack length:** 4, **discount factor:** 0.99, **learning rate:** 0.0001, **optimizer:** RMSProp (shared across all threads), **threads:** 8, **regularization:** 0.01 * policyEntropy, **policy network:** shared 4 * 512 Relu layers followed by a softmax layer for policy and a linear layer for state value function, **learning step:** 10m

Results

- Simulation running period: Jan 1st to Mar 31st (Pittsburgh winter)
- **Baseline 21.1:** heating/cooling setpoint 21.1C/23.9C for working hours
- **Baseline 23.9:** heating/cooling setpoint 23.9C/26.0C for working hours
- **Single zone testing:** using the testing environment and using the RL agent to control the *south zone*; **Multi-zone testing:** using the testing the environment and using the RL agent to control *four perimeter zones*.

Case	South Zone Mean PPD (%)	South Zone Std PPD (%)	Total HVAC Energy (kWh)
Training			
linear R λ=0.3, default	23.06	17.13	10974
linear R λ=0.4, default	20.47	15.91	11432
linear R λ=0.5, default	14.3	11.75	11940
linear R λ=0.6, default	9.95	6.38	12914
linear R λ=0.7, default	10.5	8.28	13002
l2 R, default action	12.76	9.59	12557
linear R λ=0.6, exp1	10.45	7.48	12684
linear R λ=0.6, exp2	11.56	7.17	12513
baseline 21.1	17.52	13.63	11820
baseline 23.9	13.23	12.76	12954
Single Zone Testing			
linear R λ=0.6, default	11.23	9.65	13627
linear R λ=0.6, exp1	12.29	9.58	13560
baseline 21.1	15.17	7.86	13479
baseline 23.9	7.76	4.31	14739
Multi-zone Testing			
Case	**Multi-zone Mean PPD (%)**	**Multi-zone Std PPD (%)**	**Total HVAC Energy (kWh)**
linear R λ=0.6, default	11.1	8.25	14320
baseline 21.1	14.3	7.14	13478
baseline 23.9	7.39	4.8	17237

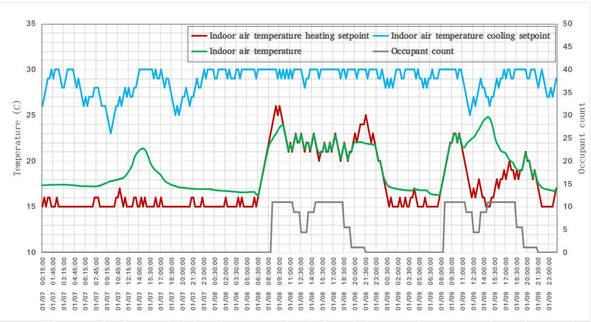

Figure 3: Control behavior snapshot of training environment (south zone)

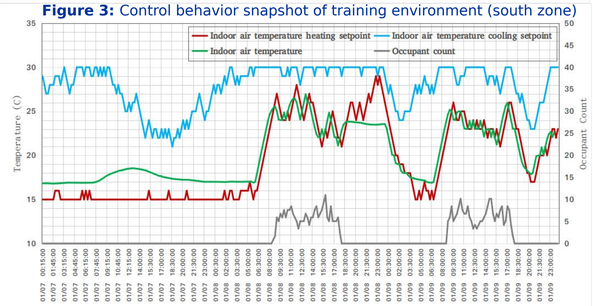

Figure 4: Control behavior snapshot of multi-zone testing environment (north zone)

Conclusion and Future Work

- Deep RL agent can achieve much better thermal comfort and some energy saving compared to the baselines in training; it can achieve 17% energy saving with still acceptable thermal comfort in the multi-zone testing even though the agent was not trained for such environment.
- Future work should focus on better state/action/reward design, multi-agent task and transfer learning.

References

[1] Lawrence Berkeley National Laboratory, EnergyPlus, 2017, http://energyplus.net

[2] Department of Energy, Commercial Reference Buildings, 2017, http://energy.gov/eere/buildings/commercial-reference-buildings

[3] Lawrence Berkeley National Laboratory, Occupancy Simulator, 2017, http://occupancysimulator.lbl.gov/

[4] Volodymyr Mnih, Mehdi Mirza, Alex Graves, Tim Harley, Timothy P Lillicrap, and David Silver. Asynchronous Methods for Deep Reinforcement Learning. In Proceedings of the 33 rd International Conference on Machine Learning, volume 48, New York, NY, USA, 2016