Towards greener, smarter and healthier office building management via cognitive digital twin

Li, Jie

2021

Li, J. (2021). Towards greener, smarter and healthier office building management via cognitive digital twin. Master's thesis, Nanyang Technological University, Singapore.
<https://hdl.handle.net/10356/149984>

<https://hdl.handle.net/10356/149984>

<https://doi.org/10.32657/10356/149984>

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Downloaded on 26 Apr 2023 15:59:31 SGT

TOWARDS GREENER, SMARTER AND HEALTHIER
OFFICE BUILDING MANAGEMENT VIA
COGNITIVE DIGITAL TWIN

Jie LI

School of Computer Science and Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfilment of the requirement for the degree of
Master of Engineering (M.Eng)

2021

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarised materials, and has not been submitted for a higher degree to any other University or Institution.

26/12/2020

.....

Date

李杰

.....

Jie LI

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

26/12/2020

.....

Date

.....

Prof. Yonggang WEN

Authorship Attribution Statement

This thesis contains material from [4 number] paper(s) published in the following peer-reviewed journal(s) / from papers accepted at conferences in which I am listed as an author.

Part of Chapter 1, Chapter 2 and Chapter 4 are published as [Guanyu Gao, Jie Li and Yonggang Wen, “DeepComfort: Energy-Efficient Thermal Comfort Control in Buildings Via Reinforcement Learning,” in IEEE Internet of Things Journal, vol. 7, no. 9, pp. 8472-8484, Sept. 2020, doi: 10.1109/JIOT.2020.2992117.](#)

The contributions of the co-authors are as follows:

- Prof. Guanyu Gao wrote and revised the manuscript.
- I co-designed the structure of experiment platform including DDPG HVAC control strategy, TRNSYS simulation environment.
- I implemented components of DDPG-based HVAC control strategy, TRNSYS-based experiment environment and data processing scripts independently.
- I conducted the experiment, analyzed the data and drew figures.
- Prof. Yonggang Wen provided the initial idea and research direction.

Part of Chapter 4 is published as [Jie Li, Wei Zhang, Guanyu Gao, Yonggang Wen, Guangyu Jin and Georgios Christopoulos “Towards Intelligent Multi-Zone Thermal Control with Multi-Agent Deep Reinforcement Learning,” in IEEE Internet of Things Journal, doi: 10.1109/JIOT.2021.3051400.](#)

The contributions of the co-authors are as follows:

- I took over this study after the previous team member left.
- On supervisor’s assignment, I redid all the experiments and prepared the manuscript draft.
- The manuscript was revised by Prof. Wei Zhang.
- Prof. Guanyu Gao assisted me with debugging of the multi-agent algorithm.
- Prof. Yonggang Wen provided the initial idea.
- Dr. Guangyu Jin and Prof. Georgios Christopoulos gave suggestions on experiment design.

Part of Chapter 1 is published as [Jie Li, Yonggang Wen, and Wei Zhang, “Poster abstract: An Urban Tropical Thermal Comfort Dataset for Smart Buildings,” in The 7th ACM International Conference on Systems for EnergyEfficient Buildings, Cities, and Transportation \(BuildSys ’20\), November 18–20, 2020, Virtual Event, Japan. ACM, New York, NY, USA, 318–319. DOI:10.1145/3408308.3431121.](#)

The contributions of the co-authors are as follows:

- I wrote the draft of the manuscript.
- Prof. Yonggang Wen provided the initial idea.
- The manuscript was revised together with Prof. Wei Zhang.

Part of Chapter 6 is published as [Wei Zhang, Jie Li, Yong Luo and Yonggang Wen, “Towards A Wearable Crowdsourcing System to Monitor Respiratory Symptoms for Pandemic Early Warning,” in IEEE Network Special Issue. Accepted.](#)

The contributions of the co-authors are as follows:

- Prof. Wei Zhang wrote and revised the manuscript.
- I designed and developed the MVP for preliminary experiment, collected cough dataset, analyzed the result, drew figures and released a crowdsourcing platform for cough data collection.
- Dr. Yong Luo proposed a novel Multi-sensor fusion algorithm.
- Prof. Yonggang Wen provided the initial idea.

26/12/2020

.....

Date

.....

Jie LI

Abstract

In this thesis, we propose an *AI-driven* digital twin solution to empower and drive “greener, smarter and healthier” buildings in the tropical environment (e.g., Singapore). Our solution aims to leverage the existing off-the-shelf hardware (e.g., sensors, wearables, etc.) and synergistically combine an emerging digital twin paradigm with the AIoT technologies, for digitalizing and optimizing the HVAC control strategy via a deep learning-based approach. We plan to optimize the HVAC system of buildings from the following three perspectives, energy efficiency, occupant comfort and behaviour study.

The first challenge is the implementation of the HVAC control policy because the changes of the thermal states in a building environment are influenced by various factors. The relationships among these influencing factors are hard to model and are always different in different building environments. To address this challenge, we propose a deep reinforcement learning based framework, **DeepComfort**, for thermal comfort control in buildings. We validate the performance of DeepComfort in simulation environment and the experimental results show that our approaches can improve the performance of thermal comfort prediction by 14.5% and reduce the energy consumption of HVAC by 4.31% while improving the occupants’ thermal comfort by 13.6%.

The second challenge we meet with is the modeling of human behaviour. We want to derive a novel human behaviour influence mechanism to motivate occupants’ behavioural changes to form sustainable energy-saving habits. Hence, we conduct national surveys, human behavioural experiments, and AI-based deep analytics to build the online personalized behaviour influence models for energy-saving behaviour studies. The national survey helps us better understand people’s orientations and perceptions of energy-saving behaviour, and we get insights in both social and psychological ways from the survey. We upgrade the HVAC system of the experiment lab, and design the new experiment

procedure for data collection and building influence model. We also design the problem formulation of our behaviour influence model. However, due to the explosion of COVID-19, our original plan was delayed because large gatherings are strictly not allowed. As a result, the experiment starts in December 2020 and it is not finished yet. In the mean time, we extend our research of smart building in the area of health. We propose a novel method for respiratory symptoms detection by using the built-in sensors of smart device, and we verify the method getting a quite accurate result on a small scale dataset we collected ourselves in office environment.

Acknowledgments

I would like to express my gratitude to all those who helped me during my study of Master.

First of all, special acknowledgment is given to my respectable supervisor Prof. Yong-gang WEN, who led me to complete the transformation from a developer to a researcher, and Prof. WEN's patient instruction and constructive suggestions benefited me a lot.

High tribute shall be paid to Prof. Wei ZHANG and Prof. Guanyu GAO who guided me to learn how to do research and how to write a good paper.

I want to express my heartfelt thanks to the students, staff and alumni in the Cloud Application and Platform (CAP) research group. The lab mates, including Dr. Weizheng HU, Dr. Yong LUO, Mr. Ruihang WANG, Mr. Meng SHEN, Mr. Deneng XIA, Mr. Yizheng Huang, Mr. Linsen DONG, and Miss Tianshu LIU together make the lab warm as we always help and listen to each other when needed, especially when we do brainstorming for some research topics. My thanks also go to the lab manager of Innovation Lab (IL) Mr. Jianxin Lim and Data Management and Analytics Graduate Lab (DMAL) Mr. Loo Kian Hock, Thomas as their organized arrangement makes it more convenient for me to access school's facilities to support my research.

Lastly my thanks would go to my beloved family, my parents, my fiancée and a special friend, a naughty cat around my living place. Without their loving considerations, great confidence and strong support, I cannot devote myself to researching these years. Your companionship makes me ever stronger to kick off every new chapter in my life.

Contents

Abstract	v
Acknowledgments	vii
List of Figures	xii
List of Tables	xvi
List of Publications	xvii
1 Introduction	1
1.1 Background and Motivation	1
1.1.1 Thermal Comfort Modeling in Buildings	1
1.1.2 Thermal Control in Buildings	2
1.1.3 HVAC Control Policy	3
1.1.4 Human Behaviour Study for Energy Saving	4
1.2 Objective of Work	5
1.3 Major Contribution	6
1.4 Summary	6
2 Learning-based Building Management: Literature Review	8
2.1 Preliminary	8
2.1.1 Thermal Comfort	8
2.1.2 ASHRAE-55	8
2.1.3 ISO 7730	9
2.1.4 Digital Twin	9
2.1.5 HVAC	9
2.1.6 Reinforcement Learning	10
2.1.7 Offline Reinforcement Learning	10

2.2	Thermal Comfort Control in Buildings	11
2.2.1	Model-based Approaches	11
2.2.2	Learning-based Approaches	12
2.3	Human Behaviour Study	13
2.4	Open-sourced Dataset	14
2.5	Summary	14
3	Experiment Platform Implementation	15
3.1	Digital Twin Platform for Training DRL Agent	15
3.1.1	Function Design	15
3.1.2	TRNSYS Project	16
3.1.3	Bridge of Building Model and Machine Learning	17
3.1.4	Online Demo	17
3.2	Platform for Behaviour Study	19
3.3	Thermal Comfort Dataset for Smart Buildings	22
3.4	Conclusion	23
4	DeepComfort: Energy-Efficient Thermal Comfort Control	25
4.1	HVAC Control System Overview	26
4.1.1	Reference System Design	26
4.1.2	Control Flow	27
4.2	System Model and Problem Formulation	28
4.2.1	Building Thermal State	28
4.2.2	Actions: Set-Points of HVAC	28
4.2.3	Reward: Thermal Comfort Prediction	30
4.2.4	Reward: Energy Consumption of HVAC	30
4.2.5	Problem Formulation	31
4.3	Algorithms for Thermal Comfort Control	31
4.3.1	Deep Neural Network for Thermal Comfort Prediction	31
4.3.2	DDPG-based Thermal Comfort Control	33
4.3.2.1	Why DDPG?	33
4.3.2.2	Training Methodology	34

4.4	Performance Evaluation	37
4.4.1	Simulation Environment	37
4.4.2	Experiment Setting	38
4.4.3	Convergence Rates under Different Settings	38
4.4.4	Performance under Different Experiment Settings	39
4.4.5	Performance Comparison with Different Methods	41
4.5	Conclusion	43
5	DeepBehaviour: Human Behaviour Study	44
5.1	Group-based Experiment	45
5.1.1	National Survey	45
5.1.2	Design of Experiments	45
5.1.3	Experiment Procedure	46
5.2	Design of Behaviour Influence Model	47
5.2.1	Human Emulator	48
5.2.2	Experiment Environment	48
5.2.3	State: Human Emulator Status	49
5.2.4	Actions: Persuasive Action	50
5.2.5	Reward: Occupants' Satisfaction Level	50
5.2.6	Problem Formulation	51
5.2.7	Algorithms for Thermal Comfort Control	51
5.3	Preliminary Result	52
5.4	Conclusion	52
6	Health Care in Building	54
6.1	Preliminary Experiment	54
6.1.1	Preliminary Result	56
6.2	Multi-sensor Fusion Approach	57
6.2.1	Multimodal Data Motivation	57
6.2.2	Algorithm Description	58
6.2.3	Platform for Data Collection	59
6.3	Conclusion	60

7 Conclusion and Future Work	61
7.1 Conclusion	61
7.2 Future Work	62
References	63

List of Figures

1.1	The overview of our iTCM system. We use AI technology to handle the data like viral signs and surrounding environment from IoT-based sensors or wearable devices.	4
1.2	Main content and organization of the thesis. There are two main parts in this dissertation, Platform and Smart Building. In the Platform section, we present the implementation of the digital-twin-based simulation platform, and next we present the concept of Smart Building in three aspects, with the goal of “greener, smarter and healthier” office building.	7
3.1	The design of TRNSYS simulation project. The main component is the building model, in addition to the model, there are several calculators for calculating heating and energy consumption.	16
3.2	The workflow of the online HVAC control demo. Users send command from the website interface, the backend server will then launch the digital twin simulation platform. All data generated by simulation platform will be sent back to front-end via websocket protocol.	18
3.3	The online demo of the HVAC control policy. We apply the trained HVAC control model in digital twin simulation platform, this demo page shows the control policy in real-time.	19
3.4	The portal for research assistant to monitor the status of the lab. This portal allows the research assistant to check the feedback status, and it allows the research assistant to get computer-based setpoint suggestions for separate zones.	20
3.5	Our in-house OEM wristband with extensive sensors and APIs and the mobile application for information management and visualization.	21

3.6	Dataset age and thermal sensation distribution.	23
4.1	The reference design of the building thermal comfort control system. The controller obtains the thermal state information from the sensors and makes control actions by adjusting the set-points of the HVAC.	26
4.2	The control flow of the building thermal comfort control system. The deep feedforward neural network is adopted for thermal comfort prediction and DRL is adopted for thermal comfort control.	27
4.3	The structure of the deep feedforward neural network for predicting thermal comfort. The inputs of the neural network include temperature, humidity, radiant temperature, air speed, metabolic rate, and clothing insulation. The output of the neural network is the predicted thermal comfort value.	32
4.4	A comparison of DQN and DDPG for thermal control. Temperature and humidity are continuous values, DDPG can be directly applied for thermal control and DQN needs the discretization of the action space.	33
4.5	The network architecture of DDPG. The actor network specifies a control action given the current thermal state and the critic network outputs an evaluation of the action generated by the actor network.	35
4.6	The implementation of the thermal comfort control simulation system. TRNSYS is adopted for simulating the building thermal environment and the HVAC, MySQL is adopted for storing sensor data and the set-points. The control diagram of building thermal comfort control simulation in TRNSYS. The thermal comfort control simulation environment simulates the energy consumption of the HVAC and building thermal states. . . .	37
4.7	The convergence rates of the thermal comfort control algorithm under different learning rates. The algorithm can achieve the best performances when the learning rate is set as 0.001 and 0.0001. The convergence rates under different hidden sizes. A larger number of neurons in the hidden layers can make the learning and convergence faster, but there will be no differences if the hidden size is larger than 512. . . .	39

4.8	The convergence rates under different discount factors. The algorithm converges faster with larger discount factors.	
	The convergence rates under different decay rates. Larger decay rates can make the algorithm converge faster.	40
4.9	The average thermal comfort values under different thermal comfort thresholds. The average thermal comfort values are close to the thresholds.	
	The distribution of the thermal comfort values under different thermal comfort thresholds. The thermal comfort value of each time slot is closely centered around the prescribed thermal comfort threshold.	40
4.10	The average cooling loads under different thermal comfort thresholds. Increasing the thermal comfort threshold can reduce energy consumption.	
	The distribution of the cooling load of the HVAC under different thermal comfort thresholds. The distribution of cooling load moves to smaller values under a larger thermal comfort threshold.	41
4.11	Comparison of different algorithms in terms of convergence rate, energy cost and thermal comfort. It can reduce the energy consumption by 4.31%, 8.95%, 9.15% compared with DQN, Q learning, and SARSA-based approaches, respectively. It improves the occupants' thermal comfort by 13.6%, 17.6%, 8.6% compared with DQN, Q Learning, and SARSA-based approaches, respectively.	42
5.1	The workflow of DRL-based algorithm. The agent will choose different persuasive actions for different people in the building, and the overall satisfaction level will be used as reward of DRL agent.	49
5.2	The training function reward varies with episode. The policy learning of each person fast converges with two episodes.	52
6.1	The screenshot of the application for preliminary experiment. The application shows the movement of smart devices using built-in sensors in a visualization way.	55
6.2	Distinct accelerometer amplitude profiles for different scenarios with data sampled in 0.1-second resolution.	56

6.3	The workflow of our proposed solution for cough detection based on the multimodal data fusion. We collect data from different aspects and apply neural networks to handle these parameters. The weighted representation of each encoder will be regarded as the output result.	58
6.4	The platform we released for cough sample collection.	59

List of Tables

3.1	Dataset Features	24
4.1	Key Notation and Definition	29

List of Publications

International Conferences

- **Jie Li**, Yonggang Wen, and Wei Zhang, “Poster abstract: An Urban Tropical Thermal Comfort Dataset for Smart Buildings,” in Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys ’20). Association for Computing Machinery, New York, NY, USA, 318–319. DOI:<https://doi.org/10.1145/3408308.3431121>. Accepted.

International Journals

- Guanyu Gao, **Jie Li** and Yonggang Wen, “DeepComfort: Energy-Efficient Thermal Comfort Control in Buildings Via Reinforcement Learning,” in IEEE Internet of Things Journal, vol. 7, no. 9, pp. 8472-8484, Sept. 2020, doi: 10.1109/JIOT.2020.2992117. Accepted.
- **Jie Li**, Wei Zhang, Guanyu Gao, Yonggang Wen, Guangyu Jin and Georgios Christopoulos “Towards Intelligent Multi-Zone Thermal Control with Multi-Agent Deep Reinforcement Learning,” in IEEE Internet of Things Journal, doi: 10.1109/JIOT.2021.3051400. Accepted.
- Wei Zhang, **Jie Li**, Yong Luo and Yonggang Wen, “Towards A Wearable Crowd-source System to Monitor Respiratory Symptoms for Pandemic Early Warning,” in IEEE Network Special Issue. Accepted.

Chapter 1

Introduction

In this chapter, I firstly present the background and motivation of our research. After that, I present the objective of this work. In the last section, I summarize the report and present the organization of the report.

1.1 Background and Motivation

With more progress in technology, the global is paying more attention to how to live comfortably, which means people are pursuing more spiritual satisfaction than before. It is predicted that the global market is growing with cost from USD 8,030.8 million to USD 42.70 billion comparing with 2018 and 2025 [1], over five times, which means this is a market with great potential. Therefore, the concept of “smart building” has aroused tremendous interests across industry and academia.

1.1.1 Thermal Comfort Modeling in Buildings

Building’s ultimate function is to serve people. The indoor environment shall be comfortable enough for people’s working and daily life. While a comfortable building environment promotes people’s well-being and happiness, a poorly maintained environment can deteriorate cognitive performance, working efficiency, and even health. It is thus important to fully understand the thermal sensation of building occupants under different conditions. The understanding enables the building managers and operators to optimize the building control towards satisfied comfort as well as other expectations like energy efficiency. Generally, such an understanding process is called thermal comfort modeling.

The modeling methodologies can be diverse and both statistics and machine learning have been applied to model thermal comfort for the past decades. Interestingly, the methodology diversity does not lead to the diversified prerequisite, and almost all the methodologies start from the same point, *data*. Predicted mean vote (PMV) [2] is possibly the most classic and widely used thermal comfort model based on six features. The model is derived based on the ASHRAE RP-884 dataset half-century ago with the data samples non-uniformly and insufficiently distributed across different regions. As a result, PMV based on the old dataset suffers from significant inaccuracy for many scenarios [3]. Despite the fast progress of the smart building research, the data sources remain limited and stagnant, and many research works have to adopt the obsolete datasets. Thereby, we address the data obsolescence and scarcity based on the Internet-of-things (IoT) technology, and we propose an IoT-based platform for data collection and then we do thermal comfort modeling on a new dataset and get better prediction accuracy.

1.1.2 Thermal Control in Buildings

In a modern building, Heating, Ventilation and Air Conditioning (HVAC) systems are the main ways to control indoor thermal conditions. Thanks to the HVAC system, people nowadays spend a significant 90% of time in buildings for various activities from working to entertaining.

The main concern of HVAC system is the high energy consumption, which makes building energy consumption account for 20%-40% of the total energy consumption [4]. Occupants may feel too cold or too hot if the set-points of HVAC system are inappropriate, although more energy is consumed. Thus, it is necessary to study how to reduce the energy consumption of the HVAC system while keeping occupants comfort, especially with the rising of electricity price and environmental pollution.

We categorize the factors which may influence thermal comfort control in buildings into three parts, namely, 1) HVAC related factors, 2) building thermal environment related factors, and 3) human subject related factors. HVAC can control the building thermal condition by adjusting set-points of air temperature and humidity, which will also consequently change the energy consumption. Building thermal environment is determined by the physical structures of the building, indoor and outdoor thermal conditions

(e.g., weather), and heat sources (e.g., bulbs, computers). The above factors determine indoor thermal condition, and human subjects can be seen as a subjective thermal comfort evaluator of the indoor thermal condition [5]. HVAC system takes appropriate control actions to meet the satisfaction of human subjects. These factors are correlated, which requires to consider them as a whole to design a proper thermal comfort control policy.

1.1.3 HVAC Control Policy

Many approaches have been proposed for building thermal comfort control and energy optimization [6]. The model-based approaches aim to model the thermal comfort control in buildings with simplified mathematical models, such as Proportional Integrate Derivative (PID) [7, 8], Model Predictive Control (MPC) [9, 10], Fuzzy Control [9, 10], Linear-Quadratic Regulator [11]. However, the complexity of the building thermal environments and the relationships among various influencing factors are hard to be precisely modeled. Different building environments may require different modelings, because it is hard to implement a generalized approach which can be directly applied in all building environments.

Some other works adopt the learning-based approaches to learn the optimal control policy, such as the reinforcement learning approach. The reinforcement learning approach can learn the optimal control policy by interactions with the thermal environment. However, the plain reinforcement learning approaches (e.g., [12], [13], [14–17]) cannot achieve high performance if the state-action space is large. Deep Reinforcement Learning (DRL) approaches (e.g., [18, 19]) have been recently adopted to address this limitation by using deep learning to learn the representation of the large state-action space. However, the methods adopted in [18, 19] require discretization of the state-action space, which will also limit the performance of the thermal control policy.

We propose a deep reinforcement learning based framework, DeepComfort, for thermal comfort control in buildings. We formulate the thermal comfort control as a cost-minimization problem by jointly considering the energy consumption of the HVAC and the occupants' thermal comfort. We first design a deep Feed-forward Neural Network (FNN) based approach for predicting the occupants' thermal comfort, and then propose a Deep Deterministic Policy Gradients (DDPG) based approach for learning the optimal thermal comfort control policy.

Fig. 1.1: The overview of our iTCM system. We use AI technology to handle the data like viral signs and surrounding environment from IoT-based sensors or wearable devices.

1.1.4 Human Behaviour Study for Energy Saving

The existing research work [20] verifies 15 different interaction approaches for human behaviour studies on thermal comfort and energy-saving habits in USA. Their experiments show that people prefer the energy-saving schedule (e.g., system automatically implement the energy-saving operations) without asking their permissions. Moreover, most of people are willing to sacrifice their own comfort to keep their partners or roommates comfortable. As such, we can see that appropriate social interaction methods have great potential to help save building energy.

Our proposed plan takes a three-stage approach. First, we would like to conduct a national survey of Singaporeans to understand their orientations and perceptions of energy-saving behaviour. Second, we plan to conduct several group-based comparison

experiments focusing on psychological factors, economic factors and social factors respectively. Finally, we will leverage DRL technique to derive the behaviour influence models based on the experiment analytical insights. In the first stage, to better understand people’s attitude and awareness of energy saving behaviour, a national survey will be conducted among Singaporeans. The survey will include online or door-to-door recruitment for both household sampling and office building sampling in order to generate a representative sample. The initial survey sample size is 1,200. Questionnaire items will measure from a range of variables, such as cognitive, affective, and behavioural. In particular, we are interested to understand individuals’ habits, routines, and personal norms related to energy-saving behaviour. By understanding these and some other psychological variables, an energy-saving intervention can be designed with a greater probability of success. In the second stage, laboratory experiment, we will measure energy use, subjective comfort, and other physiological and psychological factors of study participants. The laboratory experiment will involve five conditions. For example, we test how removing humans from being decision-makers affects subjective comfort and other parameters, and we test the effect of individual versus group-based decisions. Also, we test whether the illusion of control moderates the effect of removing humans as decision-makers.

1.2 Objective of Work

The objective of this work is to achieve the goal of “greener, smarter and healthier” buildings in the tropical environment. We divide the target into three different modules:

- **Module 1. HVAC Control Policy in Digital Twin:** In this module, we want to develop an HVAC control algorithm and verify it in TRNSYS [21] simulation environment. We build a simulation environment in TRNSYS software and connect the HVAC controller of the environment with our developed DRL-based control algorithm.
- **Module 2. Human Behaviour Study:** We want to derive a novel human behaviour influence mechanism to motivate occupants’ behavioural changes to form sustainable energy-saving habits. We conduct national surveys, human behavioural

experiments, and AI-based deep analytics to build the online personalized behaviour influence models for energy-saving behaviour studies. Thanks to the work we have done before, there are an existing data collection system (iTCM System ¹) and an ITCNN [3] prediction model to use.

- **Module 3. Health Care:** In this module, we explore more functions of smart building such as to develop an early warning system for occupants' health (e.g., respiratory symptoms detection or fatigue detection)

1.3 Major Contribution

Major contributions are listed as below:

- We design and develop an intelligent thermal comfort management system. With this system, we can collect data for further research and get set-point suggestions from the AI model. Additionally, we develop some online demo pages to present our work.
- We propose a Deep Reinforcement Learning (DRL) framework, DeepComfort, for the thermal comfort control in buildings, and we verify it in simulation environment.
- We conduct national survey and human behavioural experiments for getting insights in both social and psychological ways, and we perform AI-based analytics to build the behaviour influence models using Deep Reinforcement Learning for energy saving studies.
- We propose a novel method for respiratory symptoms detection by using accelerometer sensor in smart devices.

1.4 Summary

Fig. 1.2 presents the organization of the dissertation, more specifically, this dissertation mainly consists of two parts, they are Platform and Smart Building. Firstly, I will

¹The detailed design could be found in Fig. 1.1

Fig. 1.2: Main content and organization of the thesis. There are two main parts in this dissertation, Platform and Smart Building. In the Platform section, we present the implementation of the digital-twin-based simulation platform, and next we present the concept of Smart Building in three aspects, with the goal of “greener, smarter and healthier” office building.

introduce the implementation of our digital twin platform and the supporting software for all our experiments in Chapter 3. Next, I will explain smart buildings from three perspectives in Chapter 4, 5 and 6, including energy-efficient HVAC control, behaviour study and health care. Lastly, I will conclude this thesis.

Chapter 2

Learning-based Building Management: Literature Review

This section is organized as follows, first I will explain the preliminary one by one, next I present the literature review.

2.1 Preliminary

2.1.1 Thermal Comfort

Thermal comfort represents occupants' satisfaction level with the surrounding environment. The current general practice is to divide the satisfaction level into seven stages from cold to hot and they are cold, cool, slightly cool, neutral, slightly warm, warm, and hot respectively.

2.1.2 ASHRAE-55

Many standards had been proposed for thermal comfort problems, namely, ASHRAE-55 [22], ISO 7730 [23] and DIN EN 15251 [24].

The most widely used standard ASHRAE-55 was firstly proposed by American Society of Heating, Refrigerating, and Air-Conditioning Engineers in 1966, the purpose of this standard is to define the condition of an environment which make majority of people feel comfortable. In fact, in the past few decades, ASHRAE-55 standard was minor revised several times based on the current research and feedback from public, for example, the 2004 ASHRAE introduced adaptive method with the adoption of computer modeling, the

purpose of adaptive method is to connect indoor environment factors and outdoor weather model to better fit the natural ventilation condition. The 2010 version of ASHRAE-55 standard focused on the improvement of thermal comfort evaluation, it introduced two methods to evaluate occupants' thermal sensation.

2.1.3 ISO 7730

ISO 7730 is quite similar to ASHRAE standard, they both consider environmental factors and human factors such as indoor temperature, humidity, metabolic rate and clothing level. The differences of ISO 7730 and ASHRAE-55 are mainly reflected in cross-nation and compliance, because ISO 7730 is an international standard and it considers more on cultural, national, and geographical differences. Additionally, ASHRAE-55 is more friendly for engineers to implement because it clearly defines every term in use, equations and even sample compliance documents.

2.1.4 Digital Twin

The concept of "Digital Twin" [25] comes up as more interaction between physical world and modeled world are needed, so that the information gap between simulation and reality can be analyzed in a more digital way. The substance of digital twin is a virtual model of a process, product or service based on assumptions. The pairing of the virtual and physical worlds lead to better understand current pain points and better monitor data and system, making it easier to spot out problems even before they occur and to find out more opportunities after adjusting assumptions. Digital twin not only focuses on the current but also suggest to think retrospectively and prospectively.

2.1.5 HVAC

The main functionalities of HVAC include heating, ventilation, and air conditioning. Specifically, heating is to generate heat to raise the air temperature in the building. Ventilation is to exchange air with the outside and circulate the air within the building. Air conditioning provides cooling and humidity control. In this work, we mainly study the set-points and the energy consumption of HVAC. We do not consider the inside mechanisms of HVAC (i.e., how the set-points are achieved by HVAC via its inside

functioning). Different HVACs may have different inner mechanisms, and we design the algorithms independent of the inner mechanisms so that they can be applied in different types of HVACs.

2.1.6 Reinforcement Learning

The architecture of Reinforcement Learning (RL) is Agent - Environment - Reward. RL agent is to learn the best strategy by controlling the environment and getting reward from the Actor Critic network. Reinforcement learning has been applied in many areas for intelligent control, e.g., autonomous vehicle, video streaming [26,27], resource provisioning [28], etc. It is concerned that how the agent should take actions in a dynamic environment to maximize the overall rewards [29]. The agent first observes the current state of the environment and selects an action to take, followed by obtaining the rewards for the action. These steps will be iterated during the learning stage and the control policy will be updated until it is converged. The optimal policy can be learned during the trials without directly modeling the system dynamics, therefore, reinforcement learning is a suitable solution if the policy can be learned by the interactions with the environment and the dynamics of the system is hard to model precisely. Deep reinforcement learning [30] can significantly improve the performance compared with reinforcement learning if the dimensions of state and action are large, because deep learning has higher capacity to learn the representation of large space and tremendous data.

2.1.7 Offline Reinforcement Learning

Offline Reinforcement Learning, also known as Batch Reinforcement Learning [31], is a variant of reinforcement learning that requires the agent to learn from an existing dataset without exploration. The key feature of Batch Reinforcement Learning is that it studies from a given dataset rather than a controllable environment. More and more attention has been paid to Offline RL algorithm as it could limit the costs in some cases which require physical objects, hardware or human interactions. For example, if we train an HVAC control model in real physical buildings, people may feel uncomfortable because the agent has to explore different set-points and the set-points could be very outrageous.

Offline RL algorithm is quite related to imitation learning, because they both learn from fixed dataset. However, the idea of imitation learning is to train a model using samples collected from an “expert”, which means the dataset is supposed to be collected from an environment with excellent operations. On the other hand, Offline RL can train a policy $\pi(a|s)$ with any collected samples. Additionally, Offline RL has chances to outperform the training data, but it is difficult for imitation learning as imitation learning can only give actions from existing training samples.

2.2 Thermal Comfort Control in Buildings

The existing approaches for building thermal comfort control can be generally classified into the following two categories: 1) the model-based approaches derive the control policy by modeling the dynamics of the environment; 2) the learning-based approaches derive the control policy by learning from the interactions with the environment.

2.2.1 Model-based Approaches

Levermore *et al.* [7] and Dounis *et al.* [8] use the Proportional Integrate Derivative (PID) method for building energy management and indoor air quality control. Shepherd *et al.* [9] and Calvino [10] propose a fuzzy control method for managing building thermal conditions and energy cost. Kummert *et al.* [32] and Wang *et al.* [33] propose the optimal control method for controlling the HVAC system. Ma *et al.* [34] introduces a Model Predictive Control (MPC) based approach for controlling building cooling systems by considering thermal energy storage. Wei *et al.* [35] adopts the MPC based approach for jointly scheduling HVAC, electric vehicle and battery usage for reducing building energy consumption while keeping temperature within comfort zone. Maasoumy *et al.* [11] proposes a tracking linear-quadratic regulator for balancing human comfort and energy consumption in buildings. Oldewurtel *et al.* [36] proposes a bilinear model under stochastic uncertainty for building climate control while considering weather predictions.

The model-based approaches strive to model the dynamics of the building thermal environment with some simplified mathematical models. However, the thermal environment is affected by various factors and complicated in nature, and the changing of

the thermal condition is hard to be modeled precisely. Moreover, the model-based approaches are always designed for a specified building environment, and it is hard to derive a generalized model-based approach that is applicable in various building environments.

2.2.2 Learning-based Approaches

Barrett *et al.* [12], Li *et al.* [13] and Nikovski *et al.* [14] adopt Q learning based approaches for the HVAC control. Zenger *et al.* [15] adopt State-Action-Reward-State-Action (SARSA) for achieving the desired temperature while reducing energy consumption. Fazenda *et al.* [16] proposes a neural fitted reinforcement learning approach for learning how to schedule thermostat temperature set-points. Yu *et al.* [37] proposes a deep reinforcement learning based approach for energy management in smart homes. Dalamagkidis *et al.* [17] designs the Linear Reinforcement Learning Controller (LRLC) using linear function approximation of the state-action value function to achieve thermal comfort with minimal energy consumption. Anderson [38] proposes a robust control framework for combined Proportional Integral (PI) control and reinforcement learning control for HVAC of buildings. Wei *et al.* [18] adopts a neural network based deep Q learning method for the HVAC control. Wang *et al.* [19] adopts a Long-Short Term Memory (LSTM) recurrent neural network based reinforcement learning controller for controlling air conditioning system.

The tabular Q learning approach, SARSA, and other plain reinforcement learning approaches adopted in [12–17, 38] are not suitable for problems with large state-action spaces, partly due to that plain reinforcement learning approaches fail to achieve satisfying generalization of the value function and policy function in large spaces.

Deep Q learning [18] and LSTM based on reinforcement learning [19] can improve the performances with neural networks, which have better generalization capacity. However, the proposed approaches in these works require discretization of the state-action space, which will decrease the control precision and the performance. To fill the performance gap, we propose a DDPG-based approach for thermal comfort control in buildings. Our method also provides the capacity for customizing the thermal comfort settings. One can customize the thermal comfort threshold according to the occupants' thermal requirements to reduce energy consumption.

2.3 Human Behaviour Study

Many research experiments have been conducted for energy-saving behaviour studies. Basically, the studies could be classified into three categories, social factors, psychological factors and economy factors. For example, Huang *et al.* [20] verifies 15 different interaction approaches for human behaviour studies, and they verify that automatic energy-saving system is preferred by people, and social factors play an important role in their decision of setpoint. Cena *et al.* [39] conduct large-scale experiments in working office and found out that workers' responses to thermal comfort are largely affected by seasonality, especially in dry areas. Also, Female tend to express more discomfort feelings than male according to thermal comfort perceptions. Whitson *et al.* [40] proposes six experiments and verifies that people who lacked control may lead to illusory patterns, and it helps to explain the effect on subjective comfort level. Kaber *et al.* [41] and Baronas *et al.* [42] examine adaptive automation on human performance and situation awareness. It is helpful to know how user involvement as examined through user control can affect automation system acceptance. Djamel *et al.* [43] raises that comfort is subject to occupants feelings and its value is reflected and appreciated only when discomfort appears. As occupants' feelings and appreciation can differ a lot, there would be a gap between the thermal comfort calculated and the thermal comfort measured. Zeiler *et al.* [44] highlights the importance to integrate the response of occupants to optimize user experience even it is known that it is difficult to predict at individual level by the environment system of smart buildings. Loop in occupants response via wireless sensor as signals to control the building temperature can both minimize occupants' complaints and save energy at a certain level. Rodin Judith [45] mentions the effect of sense of control on psychological condition, hence, we plan to test whether user control is hypothesized to have an effect on user's subjective comfort level. Considering the extremely hot weather in India, Indraganti *et al.* [46] investigates the correlation between thermal comfort and effective controls in India. The comfort range concluded from the conducted experiments is slightly higher than standard and this could possibly be influenced by occupants affordability as Madhavi found the usage of A/C was less than 10% while fans account for more than 80% in the experiments.

2.4 Open-sourced Dataset

Many studies have been done to predict thermal comfort [47] [48] but few of them open-source their dataset used in the paper [49] [50] [51]. The largest open-sourced dataset could be found is ASHRAE Global Thermal Comfort Database II [52] [53] which contains 81,846 complete sets of label surrounding environment data. Specifically, some papers [54] point out directly that they want to conduct more experiments and overcome the limitation of the number of subjects for getting a more accurate result, because small scale may cause several specific variables to be over-fitting. Moreover, to overcome the problem of insufficient training data, some researchers [55] generate thermal & energy data using simulation software TRNSYS. Therefore, creating datasets for thermal comfort research will benefit all the researchers in this area.

2.5 Summary

In this chapter, I explain related terms one by one, and present the current research progress of learning-based HVAC control policy and its drawbacks, and some insights on human behaviour study from the existing studies. Lastly, we point out that researchers lack a sufficient modern dataset for smart building research. Next I will present our work to show how we solve these research problems.

Chapter 3

Experiment Platform Implementation

3.1 Digital Twin Platform for Training DRL Agent

We plan to adopt Deep Reinforcement Learning (DRL) algorithm in HVAC control, and there are three basic components in DRL technology, agent, state and reward. Before designing the agent, we need an environment for the DRL agent to provide building state. Hence, the agent could calculate reward from the status of the building.

3.1.1 Function Design

The main function of this model is to provide an environment for training algorithms. The digital twin model provides a building model close to the real world, and it simulates the building environment and HVAC systems in building. Additionally, it enables the algorithm to read the state in real time and allows the algorithm to interact with the environment, which means the algorithm can send a command in each simulation step and affect the building status. After a careful comparison, we choose TRNSYS as the digital twin supplier due to following reasons. First, it is a professional building simulation software, and it allows users to create a specific environment, customize the material of the surface and set geometric parameters. Moreover, it provides a module naming *Type-155*¹ which allows external program to interact with every simulation step. Additionally, Type-155 calls the program at each stage of life cycle, such as global parameters, first call

¹<https://sel.me.wisc.edu/trnsys/trnlib/trnsys-matlab/type155-manual.html>

Fig. 3.1: The design of TRNSYS simulation project. The main component is the building model, in addition to the model, there are several calculators for calculating heating and energy consumption.

of the simulation, last call of the simulation and all iterative calls. Users can precisely control the simulation process. Thus, we now have an opportunity to control the real-time simulation and train the HVAC control policy.

3.1.2 TRNSYS Project

The TRNSYS project is shown in Fig 3.1. Following the official instruction, I create a building project in Simulation Studio and set the basic information such as zone size, windows fraction in external walls, weather input, type of ventilation and cooling details. After that, we consult some experts in HVAC systems and fine tune the setting of ventilation and heating sources. We follow a prevalent setting of 67% air change rate per hour for the building. The working hours of the laboratory are from 9am to 6pm. Weather

data plays a vital role in our simulation, and we choose a local weather dataset, SG-Singapore-Airp-486980, available in TRNSYS to simulate the variation of the outdoor thermal environment.

3.1.3 Bridge of Building Model and Machine Learning

However, TRNSYS does not provide direct Application Programming Interface (API) to interact with Machine Learning code in Python, and the module Type-155 allows TRNSYS to communicate with Matlab code in a specific pattern. After the testing and consultation with their technical support, I find out that Type-155 module only works when we have a matched version of Matlab installed. In other words, each compiled Type-155 dynamic-link library has a matched Matlab version. Additionally, if we are using the 32bit version of TRNSYS, we have to install 32bit Matlab accordingly.

Next is the architecture of our simulation project, which allows interaction between building environment and Pytorch controller. The control mechanism of TRNSYS software is step-based, which means TRNSYS updates its internal status regularly according to our inputted time step. Specifically, each step has its life cycle, and we can process input and global parameters first. Then, we handle the state of environment and finally, we send the pre-defined command to TRNSYS software. Based on this logic, we designed the pipeline of our project by introducing an intermediate database. We associate the current status with the next action based on time series of the simulation and all information is stored in database, which means firstly we define a sequence, and in each iteration, TRNSYS sends the current state to Matlab and the value will be stored in the database with current time series. Next our Deep Reinforcement Learning agent reads the state from the database and saves control actions back to the database. So far, this is a complete life cycle of every iteration. The detailed diagram could be found in the next subsection.

3.1.4 Online Demo

In addition to the simulation workflow for DRL agent training, we provide an online demo for showcase, and the screenshot of the interface could be found in Fig. 3.3. By clicking the “Start Simulation” button, our server will receive a request and it will launch the

Fig. 3.2: The workflow of the online HVAC control demo. Users send command from the website interface, the backend server will then launch the digital twin simulation platform. All data generated by simulation platform will be sent back to front-end via websocket protocol.

simulation software accordingly, and the websocket API will send real-time data to the user interface. One thing to highlight is that the control actions and real-time state are powered by our DeepComfort HVAC control policy model. We use this demo to visually present our result of HVAC control policy in buildings.

The detailed workflow is shown in Fig. 3.2. First, we click the button on our demo website. Then, the request is received by our server. Next, the server will launch TRNSYS and Matlab application by using “pywinauto”² library, which is used for win32 application interaction by programming. With “pywinauto” library, we can easily automate Windows GUI as it allows us to send mouse and keyboard actions to windows dialogs programingly. After that, TRNSYS will send the building environmental parameters to Matlab, and then Matlab will save them into the database. Next, AI-based control algorithm will check the database periodically and provide the optimal set-point setting and save the control action into the database. Meanwhile, Matlab also checks

²<https://pywinauto.readthedocs.io/en/latest/>

Fig. 3.3: The online demo of the HVAC control policy. We apply the trained HVAC control model in digital twin simulation platform, this demo page shows the control policy in real-time.

the database periodically. Finally, once Matlab obtains the latest control decision, it will send to TRNSYS for changing the simulation environment.

3.2 Platform for Behaviour Study

In order to smooth the behaviour study, we improve the procedure of our previous experiment [56], mainly to simplify user's operations. First, we deploy all the services on the internal server, and setup the auto-startup scripts using systemctl³. Previously we had to launch the services such as tomcat container, sensor network and message queue manually before each session. Thus, the research assistant we recruited could supervise the

³<https://en.wikipedia.org/wiki/Systemd>

Fig. 3.4: The portal for research assistant to monitor the status of the lab. This portal allows the research assistant to check the feedback status, and it allows the research assistant to get computer-based setpoint suggestions for separate zones.

experiments even without any computer background. Second, we simplify the procedure. The only thing participants need to do is to submit feedback every 15 minutes, and it could reduce the chances of making mistakes. In order to standardize the experimental procedure, we conduct several rounds of pilot testing, fix all the steps and record a video for step-by-step instructions. Hence, before each session, participants play the video and follow the instructions. It frees the research assistant from repeated explanations over and over again. Additionally, we prepare an extra set of device for each participant to avoid hardware issues, and we remove extra items in the lab to avoid confusing the participants.

In the laboratory experiment, we will measure energy use, subjective comfort, and

Fig. 3.5: Our in-house OEM wristband with extensive sensors and APIs and the mobile application for information management and visualization.

other physiological and psychological factors of study participants. We conduct experiment in terms of social and psychological factors, and for the latter one we need the AI-based decision maker involved. Therefore, I develop a portal for research assistants to get the setpoint suggested by our previous developed machine learning model and monitor the status of the room, the screenshot could be found in Fig. 3.4. In each feedback window, research assistant could easily find the person who forget to submit subjective feedback via this portal, and the RA could interact with the AI-based HVAC control policy through this portal. Specifically, since there are two groups in one session simultaneously, I develop different functions for different zones because the RA has to specify the sensor set before getting suggestions from the AI, which means when we click the refresh button for Zone 1, the algorithm uses value of the first set of sensors. Otherwise, it uses the second set of sensors.

In terms of the hardware, I customize the wristband with various sensors integrated such as heart rate sensor, blood pressure sensor and SpO2 sensor, and customize the

Bluetooth software development kit (SDK) for both iOS and Android platforms, making it easier for us to collect data as all the vital signs and environmental parameters are to be measured and processed automatically. Fig. 3.5 shows the design of our application and the OEM wristband for vital data collection. Through this application, users can submit their subjective thermal sensation and additional feedback, the vital signs collection is getting done in background through Bluetooth protocol, the app will retrieve data from the wristband periodically.

3.3 Thermal Comfort Dataset for Smart Buildings

In this section I present the dataset collected through our data collection system explained in Chapter 1.

We recruited 461 participants from our university and conducted the data collection for two years (2018 and 2019) ⁴. Most of the participants are students, so the age distribution is concentrated within 20–25 as shown in Fig. 3.6(a). The gender distribution is very balanced, and we have 50.8% male and 49.2% female participants.

The collected features are listed in Table 3.1. Altogether, we have 12 features from diverse categories. Each data item starts from time and here we record hour to reflect the comfort level change throughout a day. We also extract the user profile in terms of age, weight, height, cloth insulation, and gender to offer the possibility to study personalized thermal comfort modelling. The ambient environment is monitored with a BiB set developed by SinBerBEST. We also follow a non-intrusive data collection scheme by utilizing wearables with the minimized impact of psychological fluctuation and biased subjective feedback. From the wearables, i.e., Microsoft Band 2, we are able to track the occupant’s metabolic rate, heart rate, and skin temperature. Finally, we record the corresponding thermal sensation feedback submitted by the participants. Here, we follow the ASHRAE standard by quantizing the thermal comfort into seven levels (e.g., –3 for cold, –2 for cool, –1 for slightly cool, 0 for perfectly comfortable, 1 for slightly warm, etc.). We compare the PMV-modelled comfort and our actual obtained feedback

⁴**Jie Li**, Yonggang Wen, and Wei Zhang. 2020. “Poster abstract: An Urban Tropical Thermal Comfort Dataset for Smart Buildings,” in Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys ’20). Association for Computing Machinery, New York, NY, USA, 318–319. DOI:<https://doi.org/10.1145/3408308.3431121>

Fig. 3.6: Dataset age and thermal sensation distribution.

as shown in Fig. 3.6(b). Interesting to see that PMV tends to underestimate occupant's thermal discomfort, i.e., PMV indicates slightly cool or warm even the actual feedback is significantly cold and hot. This may shed light on future comfort modeling research about the high necessity of having a latest and specialized dataset.

The data collection is conducted in sessions, each lasts for 3.5 hours. The first half-hour is for preparation tasks, i.e., participants signing the consent forms and getting familiarized with our experimental devices. Data is only collected in the rest of 3 hours and we remind the participants to submit the thermal sensation feedback every ten minutes. After each session, we reset the experimental devices to wipe out the bio-data to ensure accuracy of some parameter such as metabolic rate.

In total, we successfully collected a unique urban tropical thermal comfort dataset with 8,972 data items. The dataset captures the thermal sensation of 461 participants and the wide ranges of temperature and humidity from 19°C to 31°C and from 37% to 84%, respectively.

3.4 Conclusion

In this chapter, I develop our digital twin platform for HVAC control strategy verification, and I solve the interaction problem between TRNSYS software and the HVAC

Table 3.1: Dataset Features

#	Feature	Code	Data Source
F_1	hour	hour	mobile App
F_2	age	age	profile
F_3	weight	weight	
F_4	height	height	
F_5	clothing level	cl	
F_6	gender	gender	
F_7	air temperature	at	BiB set
F_8	relative humidity	rh	
F_9	metabolic rate	met	wearable device
F_{10}	heart rate	hr	
F_{11}	skin temperature	st	
F_{12}	actual thermal index	ati	participants

control algorithm. In addition to the HVAC control policy research, I develop a demo page for visually presenting our experiment and result. Next I present the implementation of our software platform for behaviour study including platform development and customized wristband together with Bluetooth SDK. Finally I present the detail of the dataset collected through our data collection system.

Chapter 4

DeepComfort: Energy-Efficient Thermal Comfort Control

In this work, we propose a deep learning-based framework, DeepComfort¹, for the thermal comfort control in buildings. We first design a deep feedforward neural network with Bayesian regularization for predicting the occupants' thermal comfort. The thermal comfort prediction will be used as a feedback for thermal comfort control. Then, we propose a Deep Deterministic Policy Gradients (DDPG) based approach [57] for learning thermal comfort control policy. The control variables in thermal control (e.g., temperature, humidity) are all continuous, and DDPG is a natural solution for continuous control problem, because it can avoid the discretization of the control variables and significantly improve the performances. We implement a building thermal comfort control simulation environment to evaluate the performances of our method. The main contributions of this paper are as follows:

- *Propose* a DDPG-based approach for learning the thermal comfort control policy in buildings. It can reduce the energy consumption by 4.31% and improve the thermal comfort by 13.6%. The proposed method converges faster and requires less training data.
- *Design* a deep feedforward neural network with Bayesian regularization for predicting thermal comfort. Our method can improve the prediction accuracy by 14.5% in terms of Mean Square Error (MSE).

¹Guanyu Gao, **Jie Li** and Yonggang Wen, “DeepComfort: Energy-Efficient Thermal Comfort Control in Buildings Via Reinforcement Learning.” in IEEE Internet of Things Journal, vol. 7, no. 9, pp. 8472-8484, Sept. 2020, doi: 10.1109/JIOT.2020.2992117.

Fig. 4.1: The reference design of the building thermal comfort control system. The controller obtains the thermal state information from the sensors and makes control actions by adjusting the set-points of the HVAC.

- *Implement* a building thermal comfort control simulation environment and conduct extensive experiments to evaluate the performances under different settings.

The rest of this work are organized as follows. Section 4.1 introduces the design and workflow of the thermal comfort control system. Section 4.2 presents the system model and problem formulation. Section 4.3 presents the algorithms for predicting thermal comfort and learning thermal comfort control policy. Section 4.4 evaluates the performance of the proposed method. Section 4.5 concludes this work.

4.1 HVAC Control System Overview

4.1.1 Reference System Design

The reference design of the thermal comfort control system [58, 59] is illustrated in Fig. 4.1. The system mainly consists of the following components.

Sensors: The sensors periodically measure the thermal conditions of the indoor and outdoor building environments [60], including temperature, humidity, etc. The sensors are connected with the controller via Internet of things (IoT) networks [3, 61], and the sensors will send the collected information to the controller for making thermal control decisions.

Fig. 4.2: The control flow of the building thermal comfort control system. The deep feedforward neural network is adopted for thermal comfort prediction and DRL is adopted for thermal comfort control.

Controller: The controller collects the building thermal state information from the sensors, and collects the energy consumption information from the HVAC [60]. Based on these information, the controller will make control actions by updating the set-points of the HVAC periodically according to the thermal comfort control policy.

HVAC: The HVAC will function according to the set-points updated by the controller. For instance, if the set-point of the temperature is lower than the current indoor temperature, the HVAC will start cooling until the indoor temperature matches with the set-point temperature. If the set-point temperature is higher than the current indoor temperature, the HVAC will start heating until achieving the specified indoor temperature.

4.1.2 Control Flow

We illustrate the control flow of the system in Fig. 4.2. We adopt deep feedforward neural network for predicting the occupants' thermal comfort given the current indoor thermal state, and we use DRL for making thermal comfort control decisions. The neural network based thermal comfort predictor can be trained offline using existing thermal prediction dataset. After training, the indoor building thermal state information will be input into the thermal prediction model for predicting thermal comfort. The thermal comfort prediction and the energy consumption information of the HVAC will be used for calculating the reward during each time slot. The DRL-based controller can learn the control policy by observing the rewards for taking actions on different states. We train

the DRL-based controller using the building thermal comfort control simulation system (detailed in Section 4.4.1).

4.2 System Model and Problem Formulation

In this section, we present the system models and problem formulation. We adopt a discrete-time model, where the time is denoted as $t = 0, 1, 2, \dots$. The duration of each time slot is from several minutes to one hour. The main notations used in this paper are summarized in Table 4.1.

4.2.1 Building Thermal State

The energy consumption of the HVAC is affected by both of the indoor thermal environment and the outdoor thermal environment. For the thermal state of the indoor and outdoor thermal environments, we consider the air temperature and humidity, which have the greatest influences on the energy consumption of the HVAC and the occupants' comfort. We denote the indoor air temperature and humidity at time slot t as T_t^{in} and H_t^{in} , and the outdoor air temperature and humidity at time slot t as T_t^{out} and H_t^{out} . The air temperature and humidity can be obtained by the sensors in a smart building. Therefore, we denote the state as

$$S_t = (T_t^{in}, H_t^{in}, T_t^{out}, H_t^{out}), \quad (4.1)$$

where S_t is the state at time slot t .

4.2.2 Actions: Set-Points of HVAC

The controller can change the set-points (e.g., air temperature, humidity) of the HVAC for adjusting the indoor thermal state. We denote the set-point air temperature of the HVAC at time slot t as T_t^{set} , and denote the set-point air humidity at time slot t as H_t^{set} . We denote the control action as

$$A_t = (T_t^{set}, H_t^{set}), \quad (4.2)$$

where A_t is the action taken at time slot t .

Table 4.1: Key Notation and Definition

t	the discrete time slot, $t = 0, 1, 2, \dots$
T_t^{in}	the indoor air temperature at time slot t
H_t^{in}	the indoor air humidity at time slot t
T_t^{out}	the outdoor air temperature at time slot t
H_t^{out}	the outdoor air humidity at time slot t
T_t^{set}	the set-point air temperature at time slot t
H_t^{set}	the set-point air humidity at time slot t
M_t	predicted thermal comfort value at time slot t
Φ	thermal comfort prediction model
P_t	the energy consumption of the HVAC system at time slot t
S_t	the state at time slot t
A_t	the action taken at time slot t
R_t	the received reward at time slot t
π	the thermal control policy
γ	the discount factor for the reward
D_1, D_2	the threshold for thermal comfort
$g(\cdot), h(\cdot)$	the penalty function for thermal comfort
β	the weight of the penalty for energy consumption
α_1, α_2	Bayesian hyper-parameters
n	the number of training samples
m	the number of weights in the neural network
w_j	the j -th weight in the neural network
θ^Q, θ^μ	the parameters of the critic network and the actor network
$N(t)$	the exploration noise for training
τ	the discount factor for model update

4.2.3 Reward: Thermal Comfort Prediction

Thermal comfort prediction is to quantitatively evaluate the occupants' subjective satisfaction towards the thermal state. There are six primary factors which directly affect the occupants' thermal comfort, namely, metabolic rate, clothing insulation, air temperature, mean radiance, air speed, and humidity [2]. Thus, we predict the occupants' thermal comfort value at time slot t as

$$M_t = \Phi(T_t^{in}, H_t^{in}), \quad (4.3)$$

where M_t is the predicted thermal comfort and Φ is the thermal comfort prediction model. We adopt the deep feedforward neural network-based method for predicting the occupants' thermal comfort.

4.2.4 Reward: Energy Consumption of HVAC

The HVAC will consume energy for heating, cooling, and dehumidification. The energy consumption of the HVAC during each time slot can be obtained from the smart meter. We denote the units of energy consumed by the HVAC measured in $\text{kW} \cdot \text{h}$ during time slot t as P_t , which can be obtained from the smart meter at the end of time slot t . We only consider the overall energy consumption of the HVAC during each time slot. The detailed energy consumptions for heating, cooling, dehumidification are considered as unknown.

We denote the range as $[-D, D]$, where $-D$ and D are the thresholds for thermal comfort and $-D < D$. By jointly considering the occupants' thermal comfort M_t and the energy consumption of the HVAC P_t , we calculate the reward during time slot t as the weighted sum of the two components,

$$R_t(S_t, A_t) = -\beta P_t - \begin{cases} 0 & -D < M_t < D \\ ||M_t| - |D|| & \text{otherwise} \end{cases}, \quad (4.4)$$

where R_t is the reward for time slot t , β is the weight of the energy consumption of the HVAC.

The weigh β reflects the relative importance of the energy consumption compared to the occupants' thermal comfort. The physical meaning of the equation is that the penalty

for $\frac{1}{\beta}$ units ($\text{kW} \cdot \text{h}$) of energy consumption is equal to the penalty for the deviation of the thermal comfort value from the thermal comfort thresholds by one. If the occupants' thermal comfort is more important, β should be set as a smaller value. Otherwise, β should be set as a larger value for achieving energy-efficiency.

4.2.5 Problem Formulation

The objective of DRL is to maximize the overall discount rewards from the current time slot by deriving the optimal thermal comfort control policy. The objective function can be denoted as follow,

$$\max_{\pi} \sum_{t'=0}^{\infty} \gamma^{t'} R_{t+t'}(S_{t+t'}, A_{t+t'}), \quad (4.5)$$

where γ is the discount factor.

If the precise information of the system dynamics is known, the optimal policy can be derived using value iteration or policy iteration through model-based approaches. However, it is impossible to obtain the transition probabilities in such a complex system. It motivates us to adopt the learning-based approach to learn the optimal control policy.

4.3 Algorithms for Thermal Comfort Control

In this section, we first introduce the neural network method for thermal comfort prediction. Then, we introduce the DDPG method for learning thermal control policy.

4.3.1 Deep Neural Network for Thermal Comfort Prediction

We adopt the deep feed-forward neural network for predicting thermal comfort. The structure of the neural network for predicting thermal comfort is illustrated in Fig. 4.3. The inputs of the neural network include air temperature, humidity, mean radiant temperature, air speed, metabolic rate, clothing insulation. All of these values are numerical. The hidden layer of the neural network has two layers, and the output layer has one neuron. The output of the neural network is the predicted thermal comfort value. The activation function of the hidden layer is sigmoid function, and the activation function of the output layer is a linear function.

Fig. 4.3: The structure of the deep feedforward neural network for predicting thermal comfort. The inputs of the neural network include temperature, humidity, radiant temperature, air speed, metabolic rate, and clothing insulation. The output of the neural network is the predicted thermal comfort value.

For training the neural network, the thermal comfort prediction datasets should be adopted as the training data. These datasets are labelled by the subjects for evaluating their thermal comfort levels under different thermal states. The labelled data can be noisy, due to the variations of the psychological and physiological characteristics of the occupants. To interpolate the noisy data, we adopt Bayesian regularization [62] to avoid over-fitting. The cost function for training the neural network with Bayesian regularization is to minimize the training error using the minimal weights of the neural network,

$$\phi = \alpha_1 \sum_{i=1}^n (Y_i - Y'_i)^2 + \alpha_2 \sum_{j=1}^m w_j^2, \quad (4.6)$$

where ϕ is the cost function, α_1 and α_2 are Bayesian hyperparameters for specifying the direction of the learning process to seek (i.e., minimize error or weights), n is the number of training samples, Y_i is the i -th labelled value by the subject, Y'_i is the predicted value by the neural network, m is the number of weights in the neural network, and w_j is the j -th weight.

Fig. 4.4: A comparison of DQN and DDPG for thermal control. Temperature and humidity are continuous values, DDPG can be directly applied for thermal control and DQN needs the discretization of the action space.

4.3.2 DDPG-based Thermal Comfort Control

4.3.2.1 Why DDPG?

The set-points of the HVAC, such as air temperature and humidity, are continuous. Deep Q Network (DQN) is designed for handling problems with discrete and low-dimensional action spaces [57], therefore, it cannot be directly applied to the continuous control problem. Specifically, the optimal action for a given state in DQN must be obtained by enumerating the action-value of each action,

$$a^* = \arg \max_{a \in A} Q(s, a), \quad (4.7)$$

where $Q(s, a)$ is the action-value of action a given state s , A is the set of all actions. The optimal action a^* is the action which has the largest action-value given the current state s . Since there are an infinite number of actions for continuous actions, therefore, DQN cannot be directly applied.

Compared with DQN, DDPG is designed for the continuous control problem, and we can directly obtain the set-points of the HVAC from the outputs of DDPG. We illustrate the comparisons of DQN and DDPG in Fig. 4.4. With DDPG, the network only has two outputs, namely, the set-points of air temperature and humidity. To apply DQN in thermal comfort control, we need to discretize the action space, creating a finite number

of control actions. However, if one wants to achieve finer grained discretization, it may lead to an explosion of the number of actions. For instance, the range of humidity is from 0 to 100, if we discretize with the granularity of one, it will turn into 101 possible actions. Similarly, suppose that the range of the air temperature is from 15 to 35, and there will be 200 possible actions if the granularity is 0.1. The set-points of the HVAC require one temperature value and one humidity value, and there are 20,200 possible combinations, resulting in 20,200 possible set-points of the HVAC.

As illustrated in Fig. 4.4, in the implementation of DQN, each action is one possible combination of the temperature value and the humidity value. Each output of DQN represents one action, and the value of the output is the action-value of the corresponding action. The action which has the largest action-value will be chosen as the optimal control action. The temperature value and the humidity value of the optimal control action will be the set-points of the HVAC. Therefore, the DQN network will have 20,200 outputs in the above case. It will require more training data to train the DQN network, and the discretization of the action space will also decrease the performance. The detailed performance comparisons of DQN and DDPG will be presented in Section 4.4.

4.3.2.2 Training Methodology

DDPG adopts an actor-critic framework based on Deterministic Policy Gradient (DPG) [63]. We illustrate the network architecture of DDPG in Fig. 4.5. The actor network is denoted as $A_t = \mu(S_t|\theta^\mu)$, where S_t is the thermal state and θ^μ represents the weights of the actor network and A_t represents the control action. The actor network maps the thermal state to a specific control action (i.e., the set-points of the HVAC). The critic network is denoted as $Q(S_t, A_t|\theta^Q)$, where A_t is the specified control action by the actor network and θ^Q represents the weights of the critic network. The action-value function, $Q(S_t, A_t|\theta^Q)$, describes the expected reward by taking action A_t at state S_t by following the policy. The control action specified by the actor network will be used for selecting actions during the training. After training, only the actor network is required for making thermal comfort control actions.

The training for the DDPG network is performed by interacting with the building thermal environment. The training procedure is illustrated in Algorithm 1. At the

Fig. 4.5: The network architecture of DDPG. The actor network specifies a control action given the current thermal state and the critic network outputs an evaluation of the action generated by the actor network.

beginning of each time slot t , we first obtain the current indoor and outdoor thermal state S_t , and input the thermal state into the policy network, which will output the control action. During the training, we need to explore the state space so that the policy will not converge to local optimal solutions. Therefore, we add a random noise to the obtained control action for exploration,

$$A_t = \mu(S_t | \theta^\mu) + N(t), \quad (4.8)$$

where $N(t)$ is the exploration noise and A_t is the control action added with exploration noise. In our work, we use an Ornstein-Uhlenbeck process [64] for generating the noise, $N(t)$, for exploration. Then, the control action A_t will be applied to the HVAC. At the end of time slot t , we will obtain the new thermal state S_{t+1} and calculate the overall reward R_t during the time slot. The transition (S_t, A_t, R_t, S_{t+1}) will be stored in the replay buffer B for training the policy network and the actor network.

Algorithm 1 Training Thermal Comfort Control Policy

- 1: Initialize critic network $Q(S_t, A_t|\theta^Q)$ and actor network $\mu(S_t|\theta^\mu)$ with random weights θ^Q and θ^μ
 - 2: Initialize target network $Q'(S_t, A_t|\theta^{Q'})$ and actor network $\mu'(S_t|\theta^{\mu'})$ with $\theta^{Q'} \leftarrow \theta^Q$ and $\theta^{\mu'} \leftarrow \theta^\mu$
 - 3: Initialize replay buffer B
 - 4: **for** episode = 0,1,...,M **do**
 - 5: Obtain the initial thermal state S_0
 - 6: **for** $t = 0,1,\dots,T$ **do**
 - 7: Obtain control action A_t according to Eq. (4.8)
 - 8: Update the set-points of the HVAC according to control action A_t
 - 9: Obtain new thermal state S_{t+1} and calculate reward R_t according to Eq. (4.4)
 - at the end of time slot t
 - 10: Store (S_t, A_t, R_t, S_{t+1}) into replay buffer B
 - 11: Randomly sample N transitions from replay buffer B
 - 12: Calculate the estimated reward for each sampled transition using Eq. (4.9)
 - 13: Update the critic network by minimizing the MSE over the sampled minibatch and update the actor network using the sampled policy gradient
 - 14: Update target network Q' and μ' using Eq. (4.10)
 - 15: **end for**
 - 16: **end for**
-

In each training episode, we will randomly sample N transitions from the replay buffer for training the network. For each transition $(S_i, A_i, R_i, S_{i+1}) \in N$, the estimated reward is calculated as follow,

$$R'_i = R_i + \gamma Q'(S_{i+1}, \mu'(S_{i+1}|\theta^{\mu'})|\theta^{Q'}). \quad (4.9)$$

The critic network will be updated by minimizing the MSE between the estimated reward (R'_i calculated from Eq. (4.9)) and the reward predicted by the critic network ($Q(S_i, A_i)$) over the sampled minibatch, and the actor network will be updated using the sampled policy gradient [57]. Then, the target networks will be updated using the following equations,

$$\theta^{Q'} \leftarrow \tau \theta^Q + (1 - \tau) \theta^{Q'}, \theta^{\mu'} \leftarrow \tau \theta^\mu + (1 - \tau) \theta^{\mu'}, \quad (4.10)$$

where τ is the discount factor for model update.

Fig. 4.6: The implementation of the thermal comfort control simulation system. TRNSYS is adopted for simulating the building thermal environment and the HVAC, MySQL is adopted for storing sensor data and the set-points.

The control diagram of building thermal comfort control simulation in TRNSYS. The thermal comfort control simulation environment simulates the energy consumption of the HVAC and building thermal states.

4.4 Performance Evaluation

In this section, we first present the implementation of the thermal comfort control simulation system, and then evaluate the performances of our proposed methods.

4.4.1 Simulation Environment

The thermal comfort control simulation environment is implemented to simulate the changing of the thermal state in a building and the energy consumption of the HVAC. The main components of the thermal comfort control simulation is illustrated in Fig. 4.6(a). We use TRNSYS [21] to simulate the HVAC and the thermal state in a building. The thermal comfort control algorithm and the thermal comfort prediction algorithm are implemented with PyTorch [65], which is an open-source machine learning library. The control diagram of the thermal simulation in TRNSYS is illustrated in Fig. 4.6(b). The thermal comfort control simulation environment in TRNSYS can only be programmatically controlled and accessed with Matlab (Type 155 Module [66]), therefore, the control agent which is implemented with PyTorch cannot directly interact with TRNSYS. We use MySQL [67] as the pipeline for the interactions between the control agent and TRNSYS.

4.4.2 Experiment Setting

We simulate the building environment of a laboratory which is $307\text{ }M^2$. It has 30 occupants and 40 computers with monitors. The air change rate is 0.67/hour. The weather dataset we use for simulation is SG-Singapore-Airp-486980, which is collected from Singapore. We use 10,000 hours of the simulation data in TRNSYS for training the models, and use 5,000 hours of data for testing the performances.

In our implementation, the actor network and the critic network of DDPG have two hidden layers, and each layer has 128 neurons. We use the tanh activation function and batch normalization in each layer. We adopt Adam for gradient-based optimization and the learning rate is 0.001. The discount factor τ for model update is 0.001 and the batch size is 128. The duration of each time slot is 30 minutes and each episode consists of 48 time slots. The default weight of energy cost is 0.05. The initial exploration noise scale is 0.7 and the final noise scale is 0.1. Also, the exploration noise decreases linearly over 300 episodes to 0.1. The default discount factor is 0.5.

4.4.3 Convergence Rates under Different Settings

In this subsection, we analyze the convergence rates of the thermal comfort control algorithm under different settings. The comparisons among different settings are conducted using the same data without exploration noise.

Convergence rates under different learning rates. We evaluate the convergence rates of the thermal comfort control algorithm under different learning rates, and the results are illustrated in Fig. 4.7(a). The algorithm can converge to the best performance when the learning rate is set as 0.001 and 0.0001. When the learning rate is too large (e.g., 0.01) and too small (e.g., 0.00001), the algorithm cannot converge to the optimal performances. The algorithm converges slightly faster when the learning rate is 0.0001 than 0.001.

Convergence rates under different hidden sizes. We illustrate the convergence rates of the thermal comfort control algorithm under different hidden sizes in Fig. 4.7(b). In our implementation, the actor network and the critic network both have two hidden layers. We adjust the number of neurons in the hidden layers to observe the convergence rates of the algorithm. We can observe from Fig. 4.7(b) that a larger size of the hidden layers can increase the learning speed and the convergence rate of the algorithm. Specifically,

Fig. 4.7: The convergence rates of the thermal comfort control algorithm under different learning rates. The algorithm can achieve the best performances when the learning rate is set as 0.001 and 0.0001.

The convergence rates under different hidden sizes. A larger number of neurons in the hidden layers can make the learning and convergence faster, but there will be no differences if the hidden size is larger than 512.

the convergence rates with 512 neurons and 1024 neurons are faster than the convergence rates with 256 neurons and 128 neurons. This is because more neurons in the hidden layers can improve the learning capacities. However, when the number of neurons in the hidden layers exceeds 512, the convergence rates are no different.

Convergence rates under different discount factors. We illustrate the convergence rates of the thermal comfort control algorithm under different discount factors in Fig. 4.8(a). The discount factor γ represents the weight of the future rewards. In Fig. 4.8(a), we can observe that the algorithm converges faster with larger discount factors. Meanwhile, a larger discount factor can also slightly increases the reward. When the discount factor is larger than 0.5, the differences are not significant.

Convergence rates under different decay rates. We illustrate the convergence rates of the thermal comfort control algorithm under different decay rates τ in Fig. 4.8(b). With larger decay rates, the algorithm will converge faster. However, when decay rate is larger than 0.01, the convergence rates under different decay rates are not significant. The difference between the rewards under different decay factors are not significant.

4.4.4 Performance under Different Experiment Settings

In this subsection, we evaluate the performances of the DDPG based thermal control method under different settings.

Fig. 4.8: The convergence rates under different discount factors. The algorithm converges faster with larger discount factors.

The convergence rates under different decay rates. Larger decay rates can make the algorithm converge faster.

Fig. 4.9: The average thermal comfort values under different thermal comfort thresholds. The average thermal comfort values are close to the thresholds.

The distribution of the thermal comfort values under different thermal comfort thresholds. The thermal comfort value of each time slot is closely centered around the prescribed thermal comfort threshold.

Performance under different thermal comfort thresholds. With our method one can set different thermal comfort thresholds according to the occupants' thermal comfort requirements. We evaluate the energy consumption of the HVAC and the distribution of the thermal comfort values under different thermal comfort thresholds. In Fig. 4.9(a) we illustrate the average thermal comfort values under different prescribed thermal comfort thresholds, and it can be observed that the average actual thermal comfort value measured from the building is close to the prescribed thresholds. In Fig. 4.9(b) we illustrate the distribution of the thermal comfort value of each time slot under different prescribed thermal comfort thresholds. It can be verified that the thermal comfort values in the

Fig. 4.10: The average cooling loads under different thermal comfort thresholds. Increasing the thermal comfort threshold can reduce energy consumption. The distribution of the cooling load of the HVAC under different thermal comfort thresholds. The distribution of cooling load moves to smaller values under a larger thermal comfort threshold.

indoor environment are closely centered around the prescribed thresholds. Therefore, our method can control the thermal comfort of the occupants precisely.

Fig. 4.10(a) illustrates the average cooling load of the HVAC under different thermal comfort thresholds. It will lead to less energy consumption of the HVAC if the thermal comfort threshold is set to a larger value. Therefore, if the occupants do not have stringent thermal comfort requirement, the thermal comfort threshold can be set to a larger value for energy efficiency. We illustrate the distribution of the cooling load of the HVAC under different thermal comfort thresholds in Fig. 4.10(b). We can observe that the distribution of the cooling load move towards small values when the thermal comfort threshold is larger. Therefore, adjusting the thermal comfort threshold can control the energy consumption of the HVAC.

4.4.5 Performance Comparison with Different Methods

We compare the performance of our thermal control method (DDPG) with the following baseline methods, namely, Q Learning, SARSA, and DQN. In the baselines, the temperature is discretized with the granularity of one centi-degree and the humidity is discretized

Fig. 4.11: Comparison of different algorithms in terms of convergence rate, energy cost and thermal comfort. It can reduce the energy consumption by 4.31%, 8.95%, 9.15% compared with DQN, Q learning, and SARSA-based approaches, respectively. It improves the occupants' thermal comfort by 13.6%, 17.6%, 8.6% compared with DQN, Q Learning, and SARSA-based approaches, respectively.

with the granularity of 5 percentages. We illustrate the convergences of different methods in Fig. 4.11(a). It can be observed that DDPG can achieve a faster convergence rate compared with the other baselines. This is because DDPG does not need the discretization of the action space and has fewer number of outputs of the network. Therefore, DDPG can learn thermal comfort control policy more efficiently and requires less training data. Moreover, DDPG can also achieve a higher overall reward compared with the baselines. This verifies that our method can achieve higher performances compared with the baselines. This is because Q learning and SARSA adopt the tabular methods for storing and updating the state-action values without generalization, and DQN needs the discretization of the action space.

We illustrate the average cooling load of different methods in Fig. 4.11(b). It can be observed that the average cooling load of our method is lower than the other baseline methods. Compared with DQN, Q Learning, and SARSA-based approaches, and our

proposed DDPG-based approach can reduce the energy consumption by 4.31%, 8.95%, 9.15%, respectively. Therefore, our method can achieve higher energy-efficiency. We illustrate the average thermal comfort values of different methods in Fig. 4.11(c). The average thermal comfort value of DDPG is more close to the preset threshold 0.5 and lower than the baseline methods. Compared with DQN, Q Learning, and SARSA-based approaches, our proposed DDPG-based approach can improve the occupants' thermal comfort by 13.6%, 17.6%, 8.6%, respectively. Therefore, our method can achieve a higher degree of thermal comfort for the occupants compared with the other methods while consuming less energy. The Standard Deviation (SD) of the thermal comfort value of DDPG is 0.07, compared to 0.16 of DQN, 0.49 of Q learning, 0.50 of SARSA. This means that our method has much smaller variation of thermal comfort value and the occupants can feel more comfort with our thermal comfort control method.

4.5 Conclusion

In this work, I proposed a learning based optimization framework for optimizing the occupants' thermal comfort and the energy consumption of the HVAC in buildings. I first designed a deep neural network based method with Bayesian regularization for thermal comfort prediction, and then I adopted DDPG for controlling the HVAC for optimizing the energy consumption whiling meeting the occupants' thermal comfort requirements. I implemented a building thermal comfort control simulation system using TRNSYS, and evaluated the performances under different settings. The results show that our method can achieve better prediction performances for thermal comfort prediction, and it can achieve higher thermal comfort and energy-efficiency compared with the baseline methods. In future works, we may consider using transfer learning to improve learning efficiency via the transfer of knowledge from different HVAC systems.

Chapter 5

DeepBehaviour: Human Behaviour Study

In this research, we propose to develop an online behaviour influence model, via the deep reinforcement learning (DRL) technique, and apply them for intelligent HVAC system control as well as energy-saving behaviour studies to drive the green buildings in the tropical region (e.g., Singapore). We firstly conduct a national survey to understand people's habits on energy-saving and find insights from the survey result. Then, we use the insights from our national survey and other researchers' findings in Chapter 2 literature review, and we plan to conduct the group-based experiments in terms of psychological factors, economic factors and social factors. Next we use the data collected from group-based experiment and train the behaviour influence model via Deep Reinforcement Learning. Finally we verify it in the digital twin model we developed in Chapter 3.1. The main contributions of this work are as follows:

- *Implement* a national survey in Singapore with approximately 1,200 sample size.
- *Conduct* several group-based comparison experiments.
- *Design* a Deep Reinforcement Learning neural network for training the behaviour influence model.
- *Evaluate* the performance of the behaviour influence model in our previously developed digital twin environment.

5.1 Group-based Experiment

In this section, we will introduce how we design and conduct the experiment. We plan to conduct the behaviour study from the following perspectives, including economy factors, social factors and psychological factors.

5.1.1 National Survey

The design of group-based experiment is motivated by the prior survey, so we firstly introduce the survey. The national survey is conducted by Prof. Rosenthal’s team, and they surveyed roughly 1,100 HDB residents and asked several questions about energy use at home. Questionnaire items measure several different concepts such as psychological reactance, trust in intuition, environmental beliefs, environmental personal norm, environmental responsibility and dogmatism, and use the norm activation model and the value-beliefs-norms theory. Both of those models explain environmental behaviours. Prof. Rosenthal concludes that there are two insightful questions, first, most Singaporeans say they are in control of the HVAC setpoint at home. Second, the vast majority prefer to use a remote control manually and few preferred options involving computer-based recommendations or artificial intelligence to set the temperature. On top of the survey result, we design the experiment in terms of social factors and psychological factors. We will compare group-based decisions with individual decisions, and human decisions with AI-based decisions. This is because the national survey, we find psychological reactance is positively related to preference for an AI-based temperature setting.

5.1.2 Design of Experiments

In the laboratory experiment, we measure energy use, subjective comfort, and other physiological and psychological factors of study participants. There are four participants in each four-hour session. In the sessions, we instruct them to wear devices we designed in Chapter 3.2 which is used for measuring and collecting vital signs of participants. The vital signs data and environment data will be used for our AI algorithm to give setpoint suggestions. We finally design five conditions to compare decisions in different scenarios, they are,

- 1) *RTFC* HVAC set-point is determined by AI based on environmental and occupant parameters. Participants will be informed about how the set-point is determined.
- 2) *NDVC* HVAC set-point is determined by AI, but we inform participants that the HVAC is controlled by them.
- 3) *CNSC* HVAC set-point is determined by AI, but we let participants discuss and confirm a set-point offset every 15 minutes.
- 4) *NDVC* HVAC is set by the participants. Each participant reports the preference every 15 minutes to change the temperature plus/minus several degrees or leave it the same.
- 5) *CNSC* HVAC is set by the participants. They will discuss and confirm the set-point every 15 minutes.

The difference between condition 2 and condition 4 tests whether people’s illusion of control affects their decision-making. The difference between condition 4 and condition 5 tests the effect of social factors. Condition 1, and condition 4 and 5 test how removing humans as decision-makers affects subjective comfort.

Our original plan is to conduct the experiment from March 2020, however, the plan is delayed due to the explosion of COVID-19 as we are not allowed to hold non-essential activities. Finally, with the lifting of safe distance restrictions, we start the experiment in December 2020. The experiment is still ongoing, therefore, we will share some preliminary results here, and the detailed result analytics will be presented in the next paper we are working on.

5.1.3 Experiment Procedure

There are two sessions held for each day, and each of them lasts for four hours. In order to involve more participants at the same time, we prepare two separate zones, which can contain eight people at maximum and ready to conduct different types of experiments simultaneously. In the beginning of each session, we check device battery levels to make sure that all devices are fully charged and put consent forms in place. The consent form is required by university because the privacy of all the data we collected from

participants is under the supervision and protection of the Institutional Review Board (IRB) of our university. It is also very important for us to check the instructions on desk for the participants are correct as we set different targets for each session. During each session, participants are required to read the instruction first and then start the experiment sessions. We will go into the lab and remind participants to submit the feedback every 15 minutes. Additionally, all the data generated in the experiment will be properly recorded such as the set-point input to the HVAC control unit, because all the data is necessary for training the behaviour influence model.

5.2 Design of Behaviour Influence Model

Although the experiment is not finished yet, we find some interesting findings during the sessions. The set-point determined by group-based consensus is about two degrees Celsius higher than the set-point determined by individual, which means social factors play a very important role in decision making in energy-saving. The result is exactly matching to our experiment design, because we are hoping to better understand the influence that may be exerted by economic factors, psychological factors and social factor on energy saving from the experiment. However, in terms of psychological factors, we do not find obvious discrepancy as, for some sessions, occupants' responses tend to be relatively conservative and energy-saving as they thought themselves are in control of the air condition. No matter whether occupants know that is AI to help them make the decision, individual-based decisions tend to be more aggressive than group-based, which means they tend to submit a lower set-point. Further and more accurate data analysis will be conducted after we clean all collected data. Those conclusion summarized will support the design of our behaviour influence model. We will firstly set out some persuasive actions based on the previous analysis from several different perspectives such as physiological, psychological, social and economy factors. Then, the agent of our influence model will interact with different people and select a proper action under different conditions. By the meantime, we will design a human emulator to simulate human beings, thus making our AI decision making strategy closer to the reality.

5.2.1 Human Emulator

In order to better complete the study, we plan to implement an emulator based on real human beings behaviour for evaluating the performance of our influence model. We developed a controllable digital twin model for building simulation in Chapter 4, however, the previously developed digital twin model has a few limitations. For example, it does not take occupants' status into consideration. Thus, we would like to add one more human emulator in this simulation model for training the behaviour influence model.

As states above, the main function of human emulator is to simulate the status of indoor occupants, including their biological characteristics such as gender, age, skin temperature, clothing level, and most important, personal character. Our behaviour model is mainly focusing on four perspectives, physiological, psychological, social and economic. For physiological and economic, we have obtained initial result from previous conducted study [3]. In this paper, Hu *et al.* built a prediction model ITCNN to predict thermal comfort based on vital signs and environment parameters, and from economic perspective, he raised a method using material incentive to increase the satisfaction level of occupants even under less satisfied temperature. Thus, this time, we will more focus on personality and design behaviour study from physiological and social factors.

We will allocate dozens of occupants with different personalities into our simulation environment. The personality distributions is obtained by the statistical result of our national survey and public data from the Internet. The whole idea is like this: if our behaviour model agent chooses an action aligned with occupant's personality, then he/she may have 80% chance to accept this suggestion. Otherwise, he/she may have 20% chance to accept this suggestion. If the person takes this suggestion, the way to calculate the occupant's satisfaction level may be different, which means his final satisfaction level equals to occupant's original satisfaction level add our predefined award of each action. In contrast, if the occupant does not take this suggestion, satisfaction level may decrease as his/her satisfaction level is his/her original satisfaction level minus the negative effects arising from the mismatch between the predefined action and different personality.

5.2.2 Experiment Environment

Previously, we have implemented a digital twin environment for evaluating our DRL-based HVAC control policy. Implementation of the digital twin model could be found in

Fig. 5.1: The workflow of DRL-based algorithm. The agent will choose different persuasive actions for different people in the building, and the overall satisfaction level will be used as reward of DRL agent.

Chapter 3.1, and the experiment for HVAC control policy could be found in Chapter 4. In this study, we plan to verify and evaluate the performance of the behaviour influence model in the same environment. The training workflow could be found in Fig. 5.1. DRL agent chooses different persuasive actions for different people in the building, and the overall satisfaction level will be used as reward of DRL agent.

5.2.3 State: Human Emulator Status

In the training of this behaviour influence model, we define state as the outdoor environment, indoor environment, occupants' personal status including biological and characteristic. In order to simplify the model, we can first remove all building-related factors because first, we have discussed the modeling of buildings in Chapter 4. Second, the relationship between behaviour modeling and building state is very weak. Thus, we mainly consider occupants' gender, age, skin temperature, clothing level, metabolic rate and character, and we denote these factors at time slot t as G_t , A_t , S_t , D_t , M_t and C_t respectively. Therefore, we denote the state of occupant i as O_t^i

$$O_t^i = (G_t^i, A_t^i, S_t^i, D_t^i, M_t^i, C_t^i), \quad (5.1)$$

$$S_t^{occ} = (O_t^0, O_t^1, \dots, O_t^{n-1}), \quad (5.2)$$

where S_t^{occ} is the state of occupants at time slot t . In this study, we use our previously developed digital twin model as testbed. In this model, we denote the indoor air temperature and humidity at time slot t as T_t^{in} and H_t^{in} , and the outdoor air temperature and humidity at time slot t as T_t^{out} and H_t^{out} . Therefore, the final state at time slot t can be denoted as S_t ,

$$S_t = (T_t^{in}, H_t^{in}, T_t^{out}, H_t^{out}, S_t^{occ}), \quad (5.3)$$

5.2.4 Actions: Persuasive Action

The agent of the influence model is to select a proper persuasive action under different situation for different people. We denote the persuasive action chose by the agent at time slot t for person i as P_t^i . In the environment, we denote the set-point at time slot t as T_t^{set} , and denote the humidity set-point at time slot t as H_t^{set} . Lastly, we denote the control action as

$$A_t^{occ} = (P_t^0, P_t^1, \dots, P_t^{n-1}), \quad (5.4)$$

$$A_t = (T_t^{set}, H_t^{set}, A_t^{occ}), \quad (5.5)$$

5.2.5 Reward: Occupants' Satisfaction Level

In our previous research work, we have a model for predicting occupants' thermal comfort level, this result will be regarded as the original satisfaction level. We denote the original satisfaction level of occupant i at time slot t as L_t^i . We denote the predefined award of each action as W_p , and we denote the predefined negative effects to different characters c of each action as N_p^c . Finally, we denote individual reward at time slot t of occupant i as I_t^i . The total reward is calculated as below,

$$I_t^i = L_t^i + \begin{cases} W_p & \text{accepted} \\ -N_p^c & \text{rejected} \end{cases}, \quad (5.6)$$

$$R_t(S_t, A_t) = I_t^0 + I_t^1 + \dots + I_t^{n-1}, \quad (5.7)$$

5.2.6 Problem Formulation

In this study, the target if DRL agent is to maximize the overall satisfaction level by choosing proper persuasive actions. After the training, the agent will learn an optimal policy to interact with different people. The objective function can be denoted as,

$$\max_{\pi} \sum_{t'=0}^{\infty} \gamma^{t'} R_{t+t'}(S_{t+t'}, A_{t+t'}), \quad (5.8)$$

where γ is the discount factor.

5.2.7 Algorithms for Thermal Comfort Control

The agent in Deep Reinforcement Learning can do self-learning via interaction with environment. What we need to do is to design the reward function, thus driving the agent to observe the environment, and to explore different actions to maximize reward. With the designing of the precise reward function, our agent can learn a proper controlling logic mindset. Multi-armed bandit is widely applied in many areas, including advertisement exhibitions, medical experiments, financial analysis and others. In our research, we assume that there are n people in the building and none of their responses to persuasive actions under different status is known to us. We need to give one persuasive action to each person each time to maximize their satisfaction level. In order to properly address this question, our agent needs to learn the best persuasive action not limited to single person, but also to random number of people. It is known that the agent can act differently as each persuasive action has its different chances of reward. The way out is to let our agent to adapt to different state of the environment and find the optimal action for each person, thus our agent can maximize the reward and learn to interact with all people in the building. To implement this algorithm, we use feed-forward neural network and we use a policy gradient method to update the agent. Next we train the agent by getting human state from the environment, take a persuasive action, and get the reward from each person. Then we update the network and do another iteration.

Fig. 5.2: The training function reward varies with episode. The policy learning of each person fast converges with two episodes.

5.3 Preliminary Result

In our preliminary experiment, we define three people and each one of them has different probabilities to reject or accept under different actions. After applying the contextual bandits algorithms, we draw Fig. 5.2 to demonstrate the relationship between reward and episode, with episode as x-axis and reward as y-axis respectively. In our design of experiment, the learning rate is 0.001 and the agent has 60% chance of exploring a new action, with maximum episode of 3000. Also, we sample every 300 episode. From the Fig. 5.2, we can see that both individual reward and total reward start to converge in the fourth episode, which means our agents can learn a proper policy after several interactions. In the next step, we will keep improving the human emulator and add some environmental context to make it even more complex. In addition, we will test the performance of above algorithm under different settings from different aspects, such as convergence rate and the total reward, find out the optimal experiment setting, and explain for the result.

5.4 Conclusion

In this study, I propose to develop an online personalized behaviour influence model for instilling energy-saving habits. We conduct national survey to better understand

people's attitude and awareness of energy-saving behaviour, and we get several insightful findings. Then we design group-based experiment on top of the findings in national survey and other researchers' work in psychology domain. With the preliminary result we obtain from the experiment, I design our behaviour influence model and the problem formulation. After the data collection experiment, I will further improve our model design and evaluate its performance in energy-saving and occupants' satisfaction level perspectives.

Chapter 6

Health Care in Building

COVID-19 has dramatically changed our human society. The entire world is in panic. The current impact is widespread and there is no foreseeable end. We are fighting the pandemic to restore the way forward. However, unfortunately, the effective and widely accessible vaccines are not yet available, and even the testing has not been scaled up to a satisfying level.

One promising approach is to have an early warning system, and testing should be given priority to those who have been in close contact with the coronavirus and have suspicious respiratory symptoms (e.g., cough, fever, etc.) to take advantage of insufficient testing capacity. Some countries (such as Singapore) use Bluetooth-enabled devices (such as existing smartphones and inexpensive tokens) to enhance contact tracing.

In this research, we propose using crowdsourcing methods and developing a minimal viable product (MVP) with tailored algorithms for data analytics. We choose wearable devices because it provides raw data collection and multi-sensor fusion. The MVP first leverages the emerging wearables to monitor people's vital signs (e.g., skin temperature, heartbeat rate, etc.). The data can be analyzed in a centralized or distributed fashion, and people with respiratory symptoms like coughing can be detected early and timely. Once a suspicious case is found, the MVP will immediately trigger the follow-up actions like quarantine, testing, and medical treatment.

6.1 Preliminary Experiment

To verify our proposed method, we design an application to visually track the movement of the wearable device. We design five groups of experiments to find the relationship

Fig. 6.1: The screenshot of the application for preliminary experiment. The application shows the movement of smart devices using built-in sensors in a visualization way.

between cough and different sensors. There are 3-axis movement of gyro meter, 3-axis movement of accelerometer, amplitude of gyro meter, amplitude of accelerometer and accelerometer with microphone respectively. The screenshot can be found in Fig. 6.1.

For the result of our preliminary experiment, we find that there is almost no relevance between the measurement of gyro meter and cough. As gyro meter is used to measure angular velocity of an object, a small vibration can sometimes lead to a large change in the Euler angle, but the measurement of accelerometer is quite similar to cough tremor and audio amplitude. Therefore, centering on accelerometer sensors, and then cooperating with other sensors like heartbeat rate, skin temperature sensors, we have opportunities

Fig. 6.2: Distinct accelerometer amplitude profiles for different scenarios with data sampled in 0.1-second resolution.

to propose a multi-sensor fusion solution for respiratory symptoms detection.

6.1.1 Preliminary Result

We show the accelerometer amplitude profiles under different scenarios in Fig. 6.2. As we can see, the experiment volunteer barely moves during sleep, and the accelerometer amplitude is almost zero. During normal office working hours, there are some vibrations, but more or less in the form of white noise with a very small amplitude.

The profiles become interesting in the unsteady scenarios. A mild cough can be reflected with slight spikes in the profile, and its amplitude is around twice that of a working scenario. For the time slots without cough spikes, the amplitude reflects the white noise in the normal state. A cough may be mild or severe. For the latter, we can intuitively foresee a big change in magnitude. In fact, judging from the profile of the severe cough, the spikes are many times higher than the normal state amplitudes. A severe cough also lasts longer. While a mild one is short and takes around half second, a severe one can span over one second with a huge expulsion of air.

At this stage, we might wonder whether we can detect a cough from the normal working activities with an amplitude threshold. Actually, we cannot obtain the ideal amplitude from wristband because sometimes wristbands are distant from the action source like the neck and chest. Therefore, when we cough, the vibration transmitted to the arm will be greatly reduced. In contrast, the amplitude of some scenarios can be more pronounced even if we do not want it, such as when we do boxing. In Fig. 6.2 we can see that the amplitude for boxing is even much higher than that of a severe cough.

Overall, we are able to observe that the accelerometer amplitude profiles are very different for each scenario. It is difficult to detect coughing by applying amplitude alone. Therefore, we propose another approach, multi-sensor fusion Approach.

6.2 Multi-sensor Fusion Approach

In this part, we introduce our algorithm for cough detection. We first discuss the necessity of using multimodal data. Then we describe the algorithm and the present preliminary experimental results.

6.2.1 Multimodal Data Motivation

We propose to utilize the multimodal data for cough detection. From the above part, we understand that the accelerometer amplitude profile is correlated with a cough. However, such mono-modality data is insufficient for accurate detection. Cough can lead to the profile with certain patterns like the amplitude range. However, it is uncertain that those patterns are uniquely related to the cough scenario. This is especially true that there are countless scenarios available for a wristband and it is impractical to test or even list out all of them. We specifically show the profile of speaking in Fig. 6.2, and we can see the maximum amplitude is roughly the same as a mild cough. Therefore, the introduction of the data of the other modalities becomes a logical choice to compensate for the information deficit of the accelerometer amplitude data.

When it comes to coughing, the first thing we think of is the unique sound, not the accelerometer vibrations we monitor in the wristband. On the bright side, audio data analysis is not unfamiliar to the research community and many advanced tools are

already available for scenario detection, i.e., coughing and knocking. As such, we no longer need to reinvent the wheel for audio data analytics and can take advantage of the latest technologies. However, according to our testing, it is still challenging to always accurately detect a cough, because cough detection in a noisy environment is inaccurate. Same as our thoughts on the accelerometer, audio data shall not be used alone for cough detection and we need the data of different modalities to reveal a complete picture.

Moreover, in addition to microphones and accelerometer sensors, there are many kinds of sensors built into the wearable devices, therefore, we can try to use advanced machine learning approach to fuse all the sensors and give prediction results and this approach is named multi-sensor fusion approach.

6.2.2 Algorithm Description

Fig. 6.3: The workflow of our proposed solution for cough detection based on the multi-modal data fusion. We collect data from different aspects and apply neural networks to handle these parameters. The weighted representation of each encoder will be regarded as the output result.

The main procedure of our solution is shown in Fig. 6.3. The input of the first stage consists of multiple modalities, such as the accelerometer amplitude profiles, the blood pressure, and the heart rate. Since the features of different modalities lie in diverse spaces, we employ neural networks to map the multimodal features into a common space, where a loss function based on tensor canonical correlation analysis [68] is adopted to maximize the feature correlations in the common space. At the same time, an auto-encoder is added to each modality during the space learning to maximally preserve the

information of different modalities. Since the importance of different modalities may vary in the prediction, we introduce the attention mechanism to automatically assign a weight for the mapped representation of each modality. Then, the weighted aggregation of the different mapped representations is utilized to derive an output score. And, the score can be regarded as the probability of success of the detection.

6.2.3 Platform for Data Collection

Fig. 6.4: The platform we released for cough sample collection.

To obtain enough cough samples for training the multimodal fusion algorithm, we develop a mobile application and its homepage, aiming to collect samples using crowd-sourcing method. The main function of this application is to record audio and track device movement while people cough, and we implement this module using the built-in library in Cocoa software development kit (SDK) naming AVFoundation¹ and CoreMotion². The two library allows developers to use the microphone and 3-axis sensors

¹<https://developer.apple.com/av-foundation/>

²<https://developer.apple.com/documentation/coremotion>

in iOS devices. Additionally, we support wearable devices connection for getting more features. For example, we integrate the SDK of our designed wristband and we can read skin temperature, blood pressure and SpO2 information from the wristband. In the next step, we plan to integrate more wristband SDKs, and this feature enables us to collect more valuable samples. Other than the data measured by sensors, we also design a quick questionnaire to log basic information of users, such as age, gender, symptoms and COVID-19 testing results. These samples will help us in the modeling work of our multi-sensor fusion approach.

6.3 Conclusion

In this research, we enable a cost-effective and accurate respiratory symptom monitoring system for pandemic early warning. We leverage the ever-popular pervasive sensing and wearable technologies and we design and implement a crowdsourcing-based MVP to monitor and detect respiratory symptoms. Cough detection is specifically demonstrated. We utilize multimodal data (e.g., accelerometer amplitude and microphone audio) and release a platform for cough samples collection. In the next step, with a sufficient large dataset, we can explore more on the advanced multi-sensor fusion approach.

Chapter 7

Conclusion and Future Work

7.1 Conclusion

In chapter 3, we introduce the design and implementation of our digital twin simulation model, demo webpage and the platform for our behaviour study. We develop a digital twin model for building simulation, and we solve the communication problem between digital twin model and the Machine Learning algorithm. We develop an online demo for showcase purpose, and in this portal we can observe real-time simulation and see our research result in a visualized way. In chapter 4, we propose a DRL-based framework for HVAC control. We firstly present the design and workflow of the thermal comfort control system, and we apply HVAC control algorithm named *DeepComfort* in digital twin simulation testbed for evaluating the performance. Next we explain how we model the building by defining State, Action and Reward respectively. Then we evaluate the performance of our proposed algorithm under different settings. Finally we get a result of up to 4.31% energy saving and 13.6% thermal comfort improvement. In Chapter 5, we propose a learning-based online personalized behaviour influence model and we propose a three-stage solution. Firstly, we conduct national survey to better understand people's awareness of energy-saving behaviour. With the findings from national survey, we design a group-based experiment involving five different conditions. Then we design the influence model using DRL technology and give the problem formulation. We will further improve our model design and evaluate its performance in the next couple of months as the experiment is still ongoing. In chapter 6, we present a novel approach for respiratory symptoms monitoring. After trying several different methods, we propose a

multi-sensor fusion approach, which allows the algorithm to fuse different kind of sensors for getting a higher prediction performance. To get a sufficient large dataset, we release a crowdsourcing platform to collect cough samples.

7.2 Future Work

In the future, regarding the HVAC control policy, we may consider using transfer learning to improve learning efficiency via the transfer of knowledge from different HVACs, or using the model-based reinforcement learning approaches to improve the learning efficiency of thermal comfort control. About the behaviour study, we will keep improving the model design and the emulator implementation. And we will evaluate the performance of influence model from both energy-saving perspective and occupants satisfaction level perspective. Lastly, for the respiratory symptoms monitoring system, our MVP can be augmented and developed to be a more comprehensive and advanced system, i.e., with a high-level of security, privacy, and explainability. We also plan to carry on our multimodal data collection. In the end, we believe that our system can well-equip us to be more prepared for the future respiratory related public health crisis.

References

- [1] A. M. Research, “Global smart building market size, share and industry forecast 2018-2025,” Aug 2019. [Online]. Available: <https://www.adroitmarketresearch.com/industry-reports/smart-building-market>
- [2] P. O. Fanger *et al.*, “Thermal comfort. Analysis and applications in environmental engineering.” *Thermal comfort. Analysis and applications in environmental engineering.*, 1970.
- [3] W. Hu, Y. Wen, K. Guan, G. Jin, and K. J. Tseng, “itcm: Toward learning-based thermal comfort modeling via pervasive sensing for smart buildings,” *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 4164–4177, 2018.
- [4] HKEMSD, “Hong kong energy end-use data 2020,” 2020.
- [5] S. P. Corgnati, M. Filippi, and S. Viazzo, “Perception of the thermal environment in high school and university classrooms: Subjective preferences and thermal comfort,” *Building and environment*, vol. 42, no. 2, pp. 951–959, 2007.
- [6] A. I. Dounis and C. Caraiscos, “Advanced control systems engineering for energy and comfort management in a building environment—A review,” *Renewable and Sustainable Energy Reviews*, vol. 13, no. 6-7, pp. 1246–1261, 2009.
- [7] G. J. Levermore, *Building energy management systems: An application to heating and control.* E & FN Spon, 1992.
- [8] A. I. Dounis, M. Bruant, M. Santamouris, G. Guarracino, and P. Michel, “Comparison of conventional and fuzzy control of indoor air quality in buildings,” *Journal of Intelligent & Fuzzy Systems*, vol. 4, no. 2, pp. 131–140, 1996.

REFERENCES

- [9] A. Shepherd and W. Batty, “Fuzzy control strategies to provide cost and energy efficient high quality indoor environments in buildings with high occupant densities,” *Building Services Engineering Research and Technology*, vol. 24, no. 1, pp. 35–45, 2003.
- [10] F. Calvino, M. La Gennusa, G. Rizzo, and G. Scaccianoce, “The control of indoor thermal comfort conditions: Introducing a fuzzy adaptive controller,” *Energy and buildings*, vol. 36, no. 2, pp. 97–102, 2004.
- [11] M. Maasoumy, A. Pinto, and A. Sangiovanni-Vincentelli, “Model-based hierarchical optimal control design for HVAC systems,” in *ASME 2011 Dynamic Systems and Control Conference and Bath/ASME Symposium on Fluid Power and Motion Control*. American Society of Mechanical Engineers, 2011, pp. 271–278.
- [12] E. Barrett and S. 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.
- [13] B. Li and L. Xia, “A multi-grid reinforcement learning method for energy conservation and comfort of HVAC in buildings,” in *Automation Science and Engineering (CASE), 2015 IEEE International Conference on*. IEEE, 2015, pp. 444–449.
- [14] D. Nikovski, J. Xu, and M. Nonaka, “A method for computing optimal set-point schedules for HVAC systems,” in *REHVA World Congress CLIMA’13*, 2013.
- [15] A. Zenger, J. Schmidt, and M. Krödel, “Towards the intelligent home: Using reinforcement-learning for optimal heating control,” in *Annual Conference on Artificial Intelligence*. Springer, 2013, pp. 304–307.
- [16] P. Fazenda, K. Veeramachaneni, P. Lima, and U.-M. O’Reilly, “Using reinforcement learning to optimize occupant comfort and energy usage in HVAC systems,” *Journal of Ambient Intelligence and Smart Environments*, vol. 6, no. 6, pp. 675–690, 2014.
- [17] K. Dalamagkidis, D. Kolokotsa, K. Kalaitzakis, and G. S. Stavrakakis, “Reinforcement learning for energy conservation and comfort in buildings,” *Building and environment*, vol. 42, no. 7, pp. 2686–2698, 2007.

REFERENCES

- [18] T. Wei, Y. Wang, and Q. Zhu, “Deep reinforcement learning for building HVAC control,” in *Design Automation Conference (DAC), 2017 54th ACM/EDAC/IEEE*. IEEE, 2017, pp. 1–6.
- [19] Y. Wang, K. Velswamy, and B. Huang, “A long-short term memory recurrent neural network based reinforcement learning controller for office heating ventilation and air conditioning systems,” *Processes*, vol. 5, no. 3, p. 46, 2017.
- [20] C.-C. Huang, S.-Y. Liang, B.-H. Wu, and M. W. Newman, “Reef: Exploring the design opportunity of comfort-aware eco-coaching thermostats,” in *Proceedings of the 2017 Conference on Designing Interactive Systems*, 2017, pp. 191–202.
- [21] “TRNSYS: Transient system simulation tool,” <http://www.trnsys.com>, accessed, October 2018.
- [22] A. Standard, “Standard 55-2010, thermal environmental conditions for human occupancy,” *American Society of Heating, Refrigerating and Air Conditioning Engineers*, 2010.
- [23] I. ISO, “7730: Ergonomics of the thermal environment—analytical determination and interpretation of thermal comfort using calculation of the pmv and ppd indices and local thermal comfort criteria,” *Management*, vol. 3, no. 605, p. e615, 2005.
- [24] E. DIN, “15251: 2012,” *Indoor environmental input parameters for design and assessment of energy performance of buildings addressing indoor air quality, thermal environment, lighting and acoustics*, 2012.
- [25] F. Tao, H. Zhang, A. Liu, and A. Y. Nee, “Digital twin in industry: State-of-the-art,” *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 2405–2415, 2018.
- [26] G. Gao, L. Dong, H. Zhang, Y. Wen, and W. Zeng, “Content-aware personalised rate adaptation for adaptive streaming via deep video analysis,” in *ICC 2019-2019 IEEE International Conference on Communications (ICC)*. IEEE, 2019, pp. 1–8.

REFERENCES

- [27] H. Zhang, L. Dong, G. Gao, H. Hu, Y. Wen, and K. Guan, “Deepqoe: A multi-modal learning framework for video quality of experience (qoe) prediction,” *IEEE Transactions on Multimedia*, 2020.
- [28] G. Gao, H. Hu, Y. Wen, and C. Westphal, “Resource provisioning and profit maximization for transcoding in clouds: A two-timescale approach,” *IEEE Transactions on Multimedia*, vol. 19, no. 4, pp. 836–848, 2016.
- [29] M. van Otterlo and M. Wiering, “Reinforcement learning and markov decision processes,” in *Reinforcement Learning*. Springer, 2012, pp. 3–42.
- [30] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, “Human-level control through deep reinforcement learning,” *Nature*, vol. 518, no. 7540, p. 529, 2015.
- [31] S. Lange, T. Gabel, and M. Riedmiller, “Batch reinforcement learning,” in *Reinforcement learning*. Springer, 2012, pp. 45–73.
- [32] M. Kummert, P. André, and J. Nicolas, “Optimal heating control in a passive solar commercial building,” *Solar Energy*, vol. 69, pp. 103–116, 2001.
- [33] S. Wang and X. Jin, “Model-based optimal control of VAV air-conditioning system using genetic algorithm,” *Building and Environment*, vol. 35, no. 6, pp. 471–487, 2000.
- [34] Y. Ma, F. Borrelli, B. Hencsey, B. Coffey, S. Benghea, and P. Haves, “Model predictive control for the operation of building cooling systems,” *IEEE Transactions on control systems technology*, vol. 20, no. 3, pp. 796–803, 2012.
- [35] T. Wei, Q. Zhu, and M. Maasoumy, “Co-scheduling of HVAC control, EV charging and battery usage for building energy efficiency,” in *Proceedings of the 2014 IEEE/ACM International Conference on Computer-Aided Design*. IEEE Press, 2014, pp. 191–196.

REFERENCES

- [36] F. Oldewurtel, A. Parisio, C. Jones, M. Morari, D. Gyalistras, M. Gwerder, V. Stauch, B. Lehmann, and K. Wirth, “Energy efficient building climate control using stochastic model predictive control and weather predictions,” in *Proceedings of the 2010 American control conference*, no. EPFL-CONF-169733. Ieee Service Center, 445 Hoes Lane, Po Box 1331, Piscataway, Nj 08855-1331 Usa, 2010, pp. 5100–5105.
- [37] L. Yu, W. Xie, D. Xie, Y. Zou, D. Zhang, Z. Sun, L. Zhang, Y. Zhang, and T. Jiang, “Deep reinforcement learning for smart home energy management,” *IEEE Internet of Things Journal*, 2019.
- [38] C. W. Anderson, D. Hittle, M. Kretchmar, P. Young, J. Si, A. Barto, W. Powell, and D. Wunsch, “Robust reinforcement learning for heating, ventilation, and air conditioning control of buildings,” *Handbook of Learning and Approximate Dynamic Programming*, pp. 517–534, 2004.
- [39] K. Cena and R. De Dear, “Thermal comfort and behavioural strategies in office buildings located in a hot-arid climate,” *Journal of Thermal Biology*, vol. 26, no. 4-5, pp. 409–414, 2001.
- [40] J. A. Whitson and A. D. Galinsky, “Lacking control increases illusory pattern perception,” *science*, vol. 322, no. 5898, pp. 115–117, 2008.
- [41] D. B. Kaber and M. R. Endsley, “The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task,” *Theoretical Issues in Ergonomics Science*, vol. 5, no. 2, pp. 113–153, 2004.
- [42] A.-M. K. Baronas and M. R. Louis, “Restoring a sense of control during implementation: how user involvement leads to system acceptance,” *Mis Quarterly*, pp. 111–124, 1988.
- [43] A. Djamel, L. Fayçal *et al.*, “An assessment of thermal comfort and users’“perceptions” in office buildings-case of arid areas with hot and dry climate,” *Energy Procedia*, vol. 74, pp. 243–250, 2015.

REFERENCES

- [44] W. Zeiler, D. Vissers, R. Maaijen, and G. Boxem, “Occupants’ behavioural impact on energy consumption: ‘human-in-the-loop’ comfort process control,” *Architectural Engineering and Design Management*, vol. 10, no. 1-2, pp. 108–130, 2014.
- [45] J. Rodin, “Aging and health: Effects of the sense of control,” *Science*, vol. 233, no. 4770, pp. 1271–1276, 1986.
- [46] M. Indraganti, “Behavioural adaptation and the use of environmental controls in summer for thermal comfort in apartments in india,” *Energy and buildings*, vol. 42, no. 7, pp. 1019–1025, 2010.
- [47] T. Cheung, S. Schiavon, T. Parkinson, P. Li, and G. Brager, “Analysis of the accuracy on pmv–ppd model using the ashrae global thermal comfort database ii,” *Building and Environment*, vol. 153, pp. 205–217, 2019.
- [48] M. Fletcher, D. Glew, A. Hardy, and C. Gorse, “A modified approach to metabolic rate determination for thermal comfort prediction during high metabolic rate activities,” *Building and Environment*, p. 107302, 2020.
- [49] S. Liu, S. Schiavon, H. P. Das, M. Jin, and C. J. Spanos, “Personal thermal comfort models with wearable sensors,” *Building and Environment*, vol. 162, p. 106281, 2019.
- [50] S. Munir, J. Francis, M. Quintana, N. v. Frankenberg, and M. Bergés, “Dataset: Inferring thermal comfort using body shape information utilizing depth sensors,” in *Proceedings of the 2nd Workshop on Data Acquisition To Analysis*, 2019, pp. 13–15.
- [51] J. Y. Park and Z. Nagy, “Data on the interaction between thermal comfort and building control research,” *Data in brief*, vol. 17, pp. 529–532, 2018.
- [52] R. J. De Dear, “A global database of thermal comfort field experiments,” *ASHRAE transactions*, vol. 104, p. 1141, 1998.
- [53] V. F. Ličina, T. Cheung, H. Zhang, R. De Dear, T. Parkinson, E. Arens, C. Chun, S. Schiavon, M. Luo, G. Brager *et al.*, “Development of the ashrae global thermal comfort database ii,” *Building and Environment*, vol. 142, pp. 502–512, 2018.

REFERENCES

- [54] K. Katić, R. Li, and W. Zeiler, “Machine learning algorithms applied to a prediction of personal overall thermal comfort using skin temperatures and occupants’ heating behavior,” *Applied Ergonomics*, vol. 85, p. 103078, 2020.
- [55] T. Nikolaou, I. Skias, D. Kolokotsa, and G. Stavrakakis, “Virtual building dataset for energy and indoor thermal comfort benchmarking of office buildings in greece,” *Energy and Buildings*, vol. 41, no. 12, pp. 1409–1416, 2009.
- [56] J. Li, Y. Wen, and W. Zhang, “An urban tropical thermal comfort dataset for smart buildings,” in *Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, 2020, pp. 318–319.
- [57] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, “Continuous control with deep reinforcement learning,” *arXiv preprint arXiv:1509.02971*, 2015.
- [58] M. Feldmeier and J. A. Paradiso, “Personalized HVAC control system,” in *2010 Internet of Things (IOT)*. IEEE, 2010, pp. 1–8.
- [59] A. Lifson and M. F. Taras, “Control of conditioned environment by remote sensor,” Sep. 30 2010, uS Patent App. 12/746,432.
- [60] J. Pan, R. Jain, S. Paul, T. Vu, A. Saifullah, and M. Sha, “An internet of things framework for smart energy in buildings: designs, prototype, and experiments,” *IEEE Internet of Things Journal*, vol. 2, no. 6, pp. 527–537, 2015.
- [61] D. Minoli, K. Sohraby, and B. Occhiogrosso, “IoT considerations, requirements, and architectures for smart buildings—energy optimization and next-generation building management systems,” *IEEE Internet of Things Journal*, vol. 4, no. 1, pp. 269–283, 2017.
- [62] F. D. Foresee and M. T. Hagan, “Gauss-Newton approximation to Bayesian learning,” in *Proceedings of the 1997 international joint conference on neural networks*, vol. 3. Piscataway: IEEE, 1997, pp. 1930–1935.

REFERENCES

- [63] D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller, “Deterministic policy gradient algorithms,” in *ICML*, 2014.
- [64] G. E. Uhlenbeck and L. S. Ornstein, “On the theory of the Brownian motion,” *Physical review*, vol. 36, no. 5, p. 823, 1930.
- [65] “Pytorch: An open source deep learning platform,” <https://pytorch.org>, accessed, October 2018.
- [66] “Type 155: TRNSYS-Matlab link,” <http://sel.me.wisc.edu/trnsys/trnlib/trnsys-matlab/type155-manual.html>, [Online; accessed Jun-2019].
- [67] “MySQL: the world’s most popular open source database,” <https://www.mysql.com>, accessed, October 2018.
- [68] Y. Luo, D. Tao, K. Ramamohanarao, C. Xu, and Y. Wen, “Tensor canonical correlation analysis for multi-view dimension reduction,” *IEEE transactions on Knowledge and Data Engineering*, vol. 27, no. 11, pp. 3111–3124, 2015.