



**HAL**  
open science

# Occupant behavior-based energy and comfort trade-off in smart buildings

Nour Haidar

► **To cite this version:**

Nour Haidar. Occupant behavior-based energy and comfort trade-off in smart buildings. Data Structures and Algorithms [cs.DS]. Université de La Rochelle, 2020. English. NNT : 2020LAROS032 . tel-03549223

**HAL Id: tel-03549223**

**<https://theses.hal.science/tel-03549223v1>**

Submitted on 31 Jan 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



LA ROCHELLE UNIVERSITÉ

*ÉCOLE DOCTORALE EUCLIDE*

LABORATOIRE L3i

**THÈSE** présentée par :

**Nour Haidar**

pour obtenir le grade de : **Docteur de La Rochelle Université**

Discipline : **Informatique et Applications**

Soutenue le : **17 Décembre 2020**

---

**Occupant behavior-based energy and comfort trade-off  
in smart buildings**

---

Dirigée par

Yacine Ghamri-Doudane

Alain Bouju

<b>RAPPORTEURS</b>	Marie-Pierre GLEIZES	Professeure des universités	UPS, Toulouse
	Florian DE RANGO	Associate Professor	Université de Calabre
	Zita VALE	Full Professor	ISEP-IPP, Porto
<b>EXAMINATEURS</b>	Christian INARD	Professeur des universités	La Rochelle Université
	Sidi Mohammed SENOUCI	Professeur des universités	Université de Bourgogne
	Nouredine TAMANI	Enseignant Chercheur	ISEP, Paris



To the greatest teacher I have ever had  
To the greatest person I know

To you Mohammed

## Acknowledgments

---

The Ph.D. was a challenging experience. These years had taught me a lot of things, in my professional and personal life. For sure, I will never be the person I was when I started it, and for sure, it was hard, and it could be harder without the help of the help and support to some people that I want to thank.

Firstly, I would like to express my gratitude to my advisors, Yacine Ghamri-Doudane and Alain Bouju, for giving me the opportunity to join their teams and work on a very interesting topic. I constantly benefit from his continuous support, condense and guidance throughout these four years of research allowing me to grow as a research scientist. His valuable advice are priceless for both my research and on my career choices.

I would like to thank Nouredine Tamani, one of the greatest persons I have known, for his help, collaboration and support, during all the thesis.

I would like to thank my reading committee, Pr. Zita Vale, Dr. Floriano De Rango Rivano and Pr. Marie-Pierre Gleizes, for accepting to review this dissertation. Special thanks goes also to Pr. Sidi Mohammed Senouci and Pr. Christian Inard Ahmed, for generously offering their time and good will to serve as my dissertation committee.

I am indebted to my friends for providing support and friendship that I needed during the hardest moments as well as all the fun we had together.

Last but not least, words cannot express how grateful I am to my family; my mother Fadila, my father Mustapha, my brothers: Mohamed Sami, and my sisters: Ilhem and Besma, for all of the sacrifices that you have made on my behalf. Your prayers for me was what sustained me thus far. I love you.

Words are not enough to thank my soulmate, wife and second half, Meryama for being here next to me during all these long years, and even before, and especially for giving me the best thing a father can dream of, my son Youssef Mohamed. I love you.

Thank you my brother and best friend Mohamed for your help, support and all the moment we had together. Thank you my brothers and friends Mohamed Houssam and Mohamed Fateh.





# Contents

<b>1</b>	<b>Introduction</b>	<b>21</b>
1.1	Thesis motivation: What is a Smart Building? . . . . .	21
1.2	Contributions . . . . .	22
1.3	Organization of the Dissertation . . . . .	24
<b>2</b>	<b>Smart Building Efficiency Problem and Occupant Behavior Understanding</b>	<b>27</b>
2.1	Introduction . . . . .	28
2.2	The What and Why Energy Management . . . . .	28
2.2.1	Energy Sources . . . . .	28
2.2.2	Smart Grid . . . . .	30
2.3	Building Efficiency . . . . .	31
2.4	Smart Building VS Intelligent Building . . . . .	33
2.4.1	From Building Data Collection to Intelligent Building . . . . .	33
2.4.2	Smart Building . . . . .	34
2.4.3	Building Energy Management System . . . . .	35
2.5	Building occupant behavior . . . . .	36
2.6	Thesis positioning . . . . .	37
2.7	Conclusion . . . . .	41
<b>3</b>	<b>Understanding Building Occupants' Behavior Based On Artificial Intelligence and IoT</b>	<b>43</b>
3.1	Introduction . . . . .	44
3.2	Context and Motivations . . . . .	45
3.3	Related Work . . . . .	46
3.4	Data Reduction Model Description . . . . .	48



---

3.4.1	Feature Selection-based Data Collection Period Selection Algorithm (FS-DCP) . . . . .	49
3.5	Experiments . . . . .	51
3.5.1	Dataset description . . . . .	51
3.5.2	Algorithm parameters ( $P_i, F_j, APA, MLC$ ) . . . . .	52
3.5.3	Results and discussion . . . . .	54
3.6	Summary of the Results . . . . .	59
3.7	Conclusion . . . . .	60
<b>4</b>	<b>Graph-based Occupant Behavior Modeling in Smart Building</b>	<b>63</b>
4.1	Introduction . . . . .	64
4.2	Context and Motivations . . . . .	65
4.3	Related Work . . . . .	67
4.4	Temporal Graph Model for Occupancy Behavior Prediction . . . . .	68
4.4.1	Building Movement Model . . . . .	68
4.4.2	Daytime Intervals and Frequent Occupant Movement Behavior Identification . . . . .	69
4.5	Experiments . . . . .	73
4.5.1	Dataset Description . . . . .	73
4.5.2	Similar occupant movement grouping and daytime identification . . . . .	74
4.5.3	Frequent Movement Sequence Identification . . . . .	78
4.5.4	Results . . . . .	78
4.6	Summary of the Results . . . . .	79
4.7	Conclusion . . . . .	81
<b>5</b>	<b>Occupant Behavior Prediction and Real-Time Correction-based Smart Building Energy Optimization</b>	<b>85</b>
5.1	Introduction . . . . .	86
5.2	Context and Motivations . . . . .	87
5.3	Related Work . . . . .	88
5.4	Methodology . . . . .	91
5.4.1	Occupant Movement Prediction (OMP) . . . . .	91
5.4.2	Real-time Occupancy prediction verification and Correction (ROC) . . . . .	95
5.5	Evaluation and results . . . . .	97
5.5.1	Dataset description . . . . .	97
5.5.2	Experiment description . . . . .	99

5.5.3	Result and discussion . . . . .	100
5.6	Summary of the Results . . . . .	102
5.7	Conclusion . . . . .	102
<b>6</b>	<b>Conclusion and Perspectives</b>	<b>105</b>
6.1	Contributions . . . . .	106
6.2	Perspectives . . . . .	109
	<b>Bibliography</b>	<b>114</b>



# List of Figures

2.1	From Sensor to Actuator. . . . .	35
2.2	BEMS: from data to decision. . . . .	36
2.3	Graph-based building representation. . . . .	39
3.1	The maximum of the model accuracy according to each data collection period for different ML algorithms. . . . .	55
3.2	Minimum number of features for <i>DCP</i> corresponding to the optimal MA (Fig. 3.1), for each ML algorithm. . . . .	56
3.3	The variation of the model accuracy according to the number of features. . . . .	57
3.3	(Cont.) The variation of the model accuracy according to the number of features. . . . .	58
3.4	The maximum of the model accuracy according to each data collection periods. . . . .	59
4.1	Example of building movements of an occupant. . . . .	69
4.2	Distance between two days with k-intervals. . . . .	71
4.3	Example of building. . . . .	75
4.4	Possible movements from R1 to R4. . . . .	76
4.5	Day 1 intervals decomposition. . . . .	78
4.6	Provided comfort for the four occupants in each day of the fourth week, plus the mean comfort for each occupant. . . . .	80
4.7	Daily saved energy and provided comfort. . . . .	81
5.1	TOMU algorithm unifies timestamps $T_1$ and $T_3$ to become $T_1$ for two different movements, as well as for $T_{t1}$ and $T_{t2}$ , which are unified as $T_{t1}$ . . . . .	92
5.2	Building Rooms representation. . . . .	98
5.3	Graph-based Building representation. . . . .	99
5.4	Percentage of daily saved Energy with 20% of prediction accuracy. . . . .	101

5.5	Occupants' comfort satisfaction level during the test week. . . . .	101
-----	---	-----

# List of Tables

- 3.1 Features description. . . . . 52
- 3.2 *DCP, NF, MA* for each *MLC* algorithm. . . . . 54
  
- 4.1 Time interval day groups and representative days calculation. . . . . 77

# Résumé

L'amélioration des normes de qualité de vie des bâtiments implique une consommation d'énergie plus élevée. Les occupants utilisent plus d'énergie pour alimenter leurs smartphones, ordinateurs, télévisions, etc. ce qui implique une augmentation de la consommation par habitant. De plus, la population mondiale augmente. Selon les Nations Unies [1], la croissance totale de la population mondiale est passée de 2.536 milliards en 1950 à 7.713 milliards en 2019. Cette augmentation de la population mondiale implique également une augmentation de la consommation d'énergie. L'augmentation de la consommation d'énergie des bâtiments a des conséquences dramatiques sur l'environnement; Réchauffement climatique, changements climatiques et autres impacts négatifs. La réduction de la consommation d'énergie est devenue une urgence mondiale, en particulier dans le secteur du bâtiment, car il est l'un des plus grands secteurs consommant de l'énergie. Toute cette demande croissante crée de nombreux nouveaux défis sociétaux et technologiques. Grâce aux nouvelles technologies, telles que l'Internet des objets (en anglais Internet of Things IoT), les facteurs de consommation énergétique des bâtiments peuvent être analysés en collectant des données à partir de capteurs déployés à l'intérieur comme à l'extérieur du bâtiment. Les travaux menés dans cette thèse visent à utiliser les données du bâtiment pour réduire sa consommation d'énergie. Le comportement des occupants du bâtiment devient une variable clé affectant l'efficacité énergétique des bâtiments. Par conséquent, dans cette thèse, nous nous concentrons sur la compréhension du comportement des occupants pour la minimisation énergétique dans les bâtiments.

Après un aperçu complet sur la consommation énergétique globale, de la smartification des bâtiments et de son impact sur la réduction de la consommation énergétique des bâtiments, nous concentrons nos efforts sur le développement d'une nouvelle vision sur la compréhension du comportement des occupants les bâtiments.

L'importance des données dans l'analyse du comportement des occupants nous pousse à nous interroger sur leur utilisation. La première contribution de la thèse est de se concentrer sur la sélection des types de données et de la fréquence de collecte des données

des bâtiments, pour la compréhension du comportement des occupants centrée sur les pièces du bâtiment. Ainsi, nous introduisons une approche pour sélectionner la période de collecte des données et les données pertinentes afin de pouvoir construire un modèle de prédiction du comportement des occupants du bâtiment, tout en satisfaisant les contraintes de précision. Notre approche utilise et compare cinq algorithmes d'apprentissage automatique (Machine Learning), appliqués à différentes périodes de collecte de données (de 1 minute à 60 minutes) et différents ensembles de données IoT (jusqu'à 9 différents capteurs par pièce). Les résultats de la simulation montrent que, parmi les 9 capteurs utilisés dans l'expérience collectant des données chaque minute, notre approche montre qu'un bon niveau de précision peut être obtenu en utilisant soit 8 capteurs collectant des données à un intervalle de temps de 20 min, soit 5 capteurs collectant des données à un intervalle de temps de 15 min, [2]. Comme deuxième contribution, nous proposons une approche centrée sur l'occupant pour se concentrer davantage sur le comportement de l'occupant lui-même dans un bâtiment plutôt que sur une seule pièce. Le but de cette approche est de proposer un niveau d'abstraction plus élevé pour le bâtiment visant une meilleure compréhension du comportement des occupants et de l'interconnexion des pièces. Nous utilisons une représentation basée sur les graphes temporelles des comportements des occupants pour la prédiction des comportements des occupants. L'objectif est de trouver un compromis entre la minimisation énergétique du bâtiment et la satisfaction du confort des occupants. Notre approche combine un algorithme de fouille de graphes (Graph Mining), un regroupement hiérarchique (hierarchical clustering) pour identifier les mouvements fréquents des occupants dans un intervalle de temps optimal de décomposition des jours de la semaine, et une formulation et résolution de problèmes multi-objectifs. Nous expérimentons notre approche sur un ensemble de données de 4 semaines de mouvements de 4 occupants à travers plusieurs pièces de bureaux. Les premiers résultats ont montré que notre modèle permet de minimiser jusqu'à 62.21% de la consommation énergie par rapport au fonctionnement conventionnel des systèmes HVAC, et remplit jusqu'à 94.02% du confort thermique des occupants [3]. Pour améliorer encore plus le confort de l'occupant, nous avons proposé une troisième approche qui, basée sur les



deux travaux précédents, propose de détecter les erreurs de prédiction du comportement des occupants. L'approche utilise une prédiction du comportement des occupants basée sur les graphes temporels, et une correction d'erreurs de prédiction en temps réel. Les prédictions faussement négatives (faux négatifs) peuvent causer une gêne aux occupants, c'est pourquoi nous utilisons la détection de l'état d'occupation des pièces en temps réel et la comparons à la prédiction du modèle et, si un faux négatif est détecté, le système corrige cette erreur. Cela permet de minimiser l'inconfort dans le bâtiment. Nous avons expérimenté notre approche sur des données simulées et les résultats obtenus montrent que notre modèle optimise jusqu'à 39.09% de l'énergie consommée par le système HVAC, et offre jusqu'à 99.39% de confort des occupants [4].

Nous utilisons la représentation temporelle du bâtiment et du comportement de l'occupant basée sur les graphes pour la prédiction du comportement des occupants.

**Mots-clés:** Bâtiments Intelligents, Efficacité Énergétique dans les Bâtiments, Comportement des Occupants, Optimisation du Confort des Occupants, Capteurs Intelligents, Apprentissage Automatique, Fouille de Graphes.



# Abstract

Improvements in building life quality standards imply higher energy consumption. Occupants are using more energy to power their smartphones, computers, televisions, etc. which implies an increase in per capita consumption. Furthermore, the world population is increasing. According to United Nations [1], total world population growth from 2.536 billions in 1950 to 7.713 billions in 2019. This increase in the world population also implies an increase in energy consumption. The increase in building's energy consumption has dramatic consequences on the environment; Global warming, climate changes, and other negative impacts. Reducing energy consumption has become a global emergency, especially in the buildings sector, because it represents one of the biggest energy consumer sectors. All this rising demand creates many new societal and technological challenges.

Thanks to new technologies, such as the Internet of Things (IoT), building energy consumption factors can be analyzed by collecting data from sensors deployed within and around the building. The work conducted in this dissertation aims on using the building's data to reducing its energy consumption. Building's occupant behavior becomes a key variable affecting building energy efficiency. Therefore, in this dissertation, we focus on occupant behavior understanding for building energy minimization.

After a comprehensive overview of global energy consumption, building smartification, and its impact on reducing building's energy consumption, we concentrate our efforts on the developments of a new vision of building's occupant behavior understanding.

The importance of data in the occupant's behavior analysis drives us to raise some questions about their use. The first contribution of the thesis is to focus on selecting building's data type and data collection frequency for room-centered occupant's behavior understanding. Thus, we introduce an approach to select the data collection period and the relevant data for building occupant behavior prediction model while satisfying accuracy constraints. Our approach uses and compares five machine learning algorithms, while applied to different data collection periods (from 1 minute to 60 minutes) and dif-

ferent sets of IoT data (up to 9 different sensors per room). Simulation results show that, within the 9 sensors used in the experiment collecting data each minute, our approach shows that a good level of accuracy can be obtained while using either 8 sensors collecting data at a 20-min interval, or 5 sensors collecting data at a 15-min interval, [2]. As a second contribution, we propose an occupant-centric approach to have a closer focus on the occupant behavior itself within a building instead of just a room. The aim of this approach is to propose a higher abstraction level for the building aiming at a better occupant behavior and rooms interconnection understanding. We use temporal graph-based building and occupant's behavior representation for occupant's behavior prediction. The goal is to find a trade-off between building energy minimization and occupant comfort satisfaction. Our approach combines a graph Mining algorithm, a hierarchical clustering to identify frequent occupant movements within an optimal time interval decomposition of the week days, and a multi-objective problem formulation and resolution. We experiment our approach on a 4-week data-set of 4 occupant movements among office rooms. The first results showed that our model helps minimizing energy consumption by up to 62.21% compared to the conventional functioning of HVAC systems, and fulfills up to 94.02% of occupants' thermal comfort, [3]. To improve even more the occupant's comfort, we have proposed a third approach that, based on the two previous works, proposes to detect occupant's behavior prediction errors. The approach uses temporal graph-based occupant behavior prediction and a real-time prediction error correction. False-negative predictions can cause occupant's discomfort, hence, we use real-time room occupancy state detection and compare it to model prediction and, if a false negative is detected, the system corrects this error. This allows minimizing the discomfort in the building. We experimented our approach on simulated data and the obtained results show that our model optimizes the HVAC energy consumption by up to 39.09%, and provides up to 99.39% of occupants comfort [4].

**Keywords:** Smart building, Building energy efficiency, Occupant behavior, Occupant comfort optimization, Smart Sensor, Machine Learning, Graph mining.



# Chapter 1

## Introduction

### Contents

---

<b>1.1 Thesis motivation: What is a Smart Building? . . . . .</b>	<b>21</b>
<b>1.2 Contributions . . . . .</b>	<b>22</b>
<b>1.3 Organization of the Dissertation . . . . .</b>	<b>24</b>

---

### 1.1 Thesis motivation: What is a Smart Building?

Buildings need to be comfortable, safe, energy-efficient, secure and meet higher occupant expectations. This can be achieved thanks to digital technologies and the Internet of Things (IoT) in particular. To attract and retain occupants, buildings need to go beyond the traditional focus of operational efficiencies, to create people-centric environments that enhance occupant experience, instead of creating a non-smart but automated building. To be called smart, a building must succeed in capturing occupant engagement, transforming the building into an extension of the people who live in or visit it. Thus, the building becomes an active contributor to both, the success of energy optimization as well as the occupant's productivity and well-being goals. Mobile and computer applications give occupants direct control over their environment, such as the possibility

of personalizing temperature and control light. These capabilities provide an occupant-centric approach and, effectively connecting the building's occupant with the building. Smart buildings consider these capabilities but need to go beyond that.

In a smart building, sensors need to be deployed within and around the building. These sensors need to be connected to state-of-the-art and new software, if necessary. The goal is to collect data from all building's subsystems into a common data lake for the whole ecosystem. Data is then analyzed and presented in a dashboard. This last offers a clear picture of how the entire building is working. It allows to occupant to interact with the building, receives notifications, such as security intrusion or indoor air quality problems. These features become common nowadays but, again, smart buildings need to go beyond that.

Indeed, the increase in the building's energy consumption represents a serious problem that can lead to the longest and irreversible environmental disasters. Building energy consumption needs to be carefully studied to be effectively minimized.

## 1.2 Contributions

In this dissertation, we aim to address the topic of occupant behavior in smart buildings. As shown in Section 1.1, the occupant represents the center of Smart Building conceptualization. He/she is able to display all buildings collected or analyzed data and then, interact with the building. It is for these reasons that we focus, in this thesis, on the occupant. However, to study the building's occupant behavior we need to have data collected about this. These are by sensors deployed within and around the building. Hence, in this thesis, we make data-based occupant behavior understanding our main topics. Accordingly, in this dissertation, we deal with the following questions: *what data to use?*, *when to collect these data?*, *how to model occupant's behavior?*, *how to use occupant behavior prediction to minimize building energy consumption and maximize occupant's comfort?*, and *how to correct occupant behavior prediction errors impacting occupant's comfort?*.

Our first contribution focuses on the two first questions, i.e. *what data to use?* and *when to collect these data?*. Data collection requires deploying sensors within and around buildings. However, this deployment costs time and money. Therefore, we need to minimize the number of sensors to use for meeting a specific need. Additionally, for many sensor types such as temperature, humidity, luminosity, etc., we need to specify the frequency of data collection as, nevertheless, the shorter the data collection frequency is, the more the sensors consume energy, thus we need to find the adequate frequency to reach our goal. In this contribution, we introduce a method for finding the equilibrium between the number of sensors (i.e. data types) to use and the data collection frequency. We aim at minimizing the number of sensors while maximizing the frequency of data collection. We use feature selection and compare many Machine Learning algorithms (Random Forest Classifier, Decision Tree Classifier, Extra-Trees Classifier, Gaussian Naive Bays, and Multi-Layer Perceptron Classifier) to reach such equilibrium and select the one that best meets our expectations. The experiments were carried out on a real dataset from one-year data collected from many rooms in E.ON Energy Research Center in Aachen Germany. Results show that it is possible to build an occupancy predictive model with the Random Forest Classifier algorithm having an accuracy of at least 90%, by using 8 sensors collecting data at a 20-min interval, or 5 sensors collecting data at a 15-min interval.

In our second contribution, we aim to answer the two following questions: *how to model occupant behavior?* and *how to use occupant behavior prediction to minimize building energy consumption and maximize occupant's comfort?*. To achieve this, we propose a graph-based building and occupant-movements representation in which building rooms are represented by graph nodes, and possible direct movements between the rooms are represented by graph edges. This representation gives a macroscopic vision of the occupant behavior within the building and the inter-connectivity between rooms based on this behavior. Furthermore, virtual rooms can be created on the graph to study a future possible physical decomposition of rooms. This occupant-centric method allows real-time occupant tracking, thus produce an accurate prediction model. We use Graph Mining algorithm to extract the prediction model. The approach combines a graph learning



algorithm, a hierarchical clustering to identify frequent occupant movements within the optimal time interval decomposition of days, and a multi-objective problem resolution. We experimented our approach on a 4-week dataset of 4 occupant movements among of-fice rooms. The first results showed that our model helps minimize energy consumption by up to 62.21% compared to conventional functioning of HVAC systems, and fulfills up to 94.02% of occupants' thermal comfort.

The third contribution aims to solve possible prediction errors of the second contribution. To do so, we introduce real-time room occupancy states to our previous verify the correctness of the prediction model. If a false-negative prediction error is detected, the system recovers the decision error by, for instant, activating a device that was wrongly turned-off. Since prediction models are not always that accurate, it is possible to face situations where HVAC of some rooms are activated while these are empty or vice-versa, leading to either a waste of energy or a lack of occupant's comfort. To deal with this issue, we make use of sensors to detect real-time occupancy of building rooms and then correct the prediction when necessary. To achieve this, we developed a graph mining-based optimization approach that combines occupant behavior prediction and a real-time correction. We experimented our approach on simulated data and results showed that our model optimizes up to 39.09% of HVAC energy consumption, and provides up to 99.39% of occupants' comfort.

### **1.3 Organization of the Dissertation**

This previous two sections discussed the importance of Building Smartification and the way such smartification could be achieved with an aim of transforming traditional buildings into modern interactive buildings by involving new technologies, the Internet of Things, and Data Science in order to put the occupants at the heart of Building's transformation. In order to deeply present our proposed approach and solution, the organization of this dissertation is as follows.

In Chapter 2, we detail the problem of energy consumption and world global warming. We also review the invention of the Smart Grid concept as well as the associated buildings' energy efficiency problem. Then, we discuss occupant behavior and its impact on building energy consumption, and the problem of finding a good compromise between building energy minimization and occupant's comfort assurance. Finally, we position our work in relation to the existing literature.

Chapter 3 details our first contribution. We introduce our approach of data types (i.e. required sensors) minimization and data collection frequency maximization. Thereafter, we list the experimental details. We finish the chapter by discussing the obtained results, then present some concluding remarks.

The Chapter 4 of the dissertation represents our second contribution. Mainly, we define a graph-based building and occupant behavior abstraction representation. We show, based on such representation, graph mining techniques allow identifying the occupant behavior within a building and thus help minimizing the energy consumption of the building while maintaining a good comfort level. Then, we explain the experiment used to validate such an approach. Thereafter, we discuss the obtained results and conclude by analyzing our findings.

Our last contribution is presented in chapter 5. We highlight a limitation identification in chapter 4, i.e. the impact of prediction errors. Then, we introduce the real-time occupant behavior prediction verification and errors correction as an add-on to our solution in chapter 4. Then, we detail the experimented study and discuss the results improvements. A conclusion about these results is given here as well.

To conclude, we summarize all dissertation contributions in the final chapter (chapter 6). The chapter also presents possible future directions that can be followed to further improve our work.



## Chapter 2

# Smart Building Efficiency Problem and Occupant Behavior Understanding

### Contents

---

<b>2.1</b>	<b>Introduction</b>	<b>28</b>
<b>2.2</b>	<b>The What and Why Energy Management</b>	<b>28</b>
2.2.1	Energy Sources	28
2.2.2	Smart Grid	30
<b>2.3</b>	<b>Building Efficiency</b>	<b>31</b>
<b>2.4</b>	<b>Smart Building VS Intelligent Building</b>	<b>33</b>
2.4.1	From Building Data Collection to Intelligent Building	33
2.4.2	Smart Building	34
2.4.3	Building Energy Management System	35
<b>2.5</b>	<b>Building occupant behavior</b>	<b>36</b>
<b>2.6</b>	<b>Thesis positioning</b>	<b>37</b>
<b>2.7</b>	<b>Conclusion</b>	<b>41</b>

---

## 2.1 Introduction

Energy is everywhere. Nowadays, we cannot leave without it. It is in our houses, cars, workplace, even in our pockets. The modern human cannot imagine his life without energy and, in turn, his need for energy keeps growing. Furthermore, according to the United Nations [5] the world population has grown from 2.6 billion in 1950 to 7.7 billion humans in 2020, and could peak at nearly 11 billion around 2100. Therefore, world energy needs are constantly increasing. Consequently, it is estimated that energy consumption in the world will increase by 41% from 2015 to 2040 [6]. Nevertheless, this increase in energy consumption is not without consequences, on the one hand, on the available energy resources and, on the other hand, on the planet.

In this chapter, Section 2.2 presents the different energy sources and the role of Smart Grid for building energy supply. In section 2.3, we discuss the building's energy efficiency problem. Section 2.4 covers the importance of data in understanding building energy performance and shows the difference between Intelligent Buildings and Smart Buildings. Building Energy Management Systems are presented in Section ???. The role of the occupant and its impact on building energy performance is presented in Section 2.5. Section 2.6 presents the problem position in relation with the studies literature. We conclude the chapter by a conclusion (Section 2.7).

## 2.2 The What and Why Energy Management

### 2.2.1 Energy Sources

The field of energy development focused on investigating new sources of energy from natural resources. Based on the origin of energy, we can distinguish two main types of energy:

- Conventional energy (also called non-renewable energy): It consists of static stores of energy extracted from underground by human interaction [7]. It represents

energy that will run out, which means that it will disappear over time, and the required time for these types of energy to be created is too long (millions of years) in comparison with the speed of its consumption.

The main types of conventional energy are Coal, Oil (also known as petroleum), Natural gas, and Nuclear. Conventional energy is a major source of energy for a vast amount of industries [8]. According to the U.S. Energy Information Administration (EIA) [9], within the U.S. primary energy consumption in 2019, around 88% comes from conventional energy (Oil: 37%, Natural gas: 32%, Coal: 11%, and Nuclear: 8%).

This source of energy has the reputation of being flexible, short term available, cheap to produce [10]. Nevertheless, the use of these energies is not without negative consequences. In fact, the production of energy based on these sources produces an enormous amount of carbon dioxide  $CO_2$ . These emissions are the main cause of global warming [11, 12, 13]. Furthermore, its transportation can lead to natural disasters, like the Japanese-owned Tanker oil spill catastrophe in Mauritius, on August 6th, 2020 [14].

- Renewable energy: It is the energy collected from natural repetitive local sources [7]. Some of these energy sources can be daily renewable. The reputation, benefits, and utilization of renewable energy is increasing all around the world.

The main types of renewable energy are Biomass, Hydro, Tidal, Wind, Solar, and Geothermal. This type of energy is less used than conventional energy. For example, in the United States of America, in 2019, renewable energy represents around 12% of the primary energy consumption, according to the U.S. Energy Information Administration (EIA) [9]. Within this 12%, Biomass represents 43% of the global renewable energy consumption (20% wood, 20% bio-fuels, and 4% biomass waste), 24% wind, 22% hydroelectric (Hydro and Tidal), 9% solar, and 2% geothermal.

Renewable energy has a good reputation for helping sustainable development, less harm to the planet (less  $CO_2$  and Greenhouse Gas effect), no air pollutant

emissions, low maintenance requirements, and it can be generated locally [7, 15]. Nonetheless, there are still some drawbacks. For this matter, we will give the disadvantages of each source of renewable energies [15]: Bad practices of Biomass exploitation can generate high greenhouse gas, air pollution can be led by open burning, and physical footprint can be left. For Wind sources, power supply fluctuation, bird kills, and noises can be seen. The problem of hydro-power is that it can affect the fish population if it is not properly designed, and, because of the irregularity of the wind, power supply fluctuation can be observed. Solar power needs storage units (likes batteries), and has a physical footprint. Geothermal may cause toxic gases to release from below the earth's surface.

An important challenge of the first half of the 21st century is to do the transition from conventional to renewable energy [16]. A trade-off between these two types of energy may be a good start to compensate for the disadvantages of each of them [17, 18, 8, 15].

### **2.2.2 Smart Grid**

As in a traditional system, energy is delivered by centralized energy providers. Extensions to these need to be added to integrate the new decentralized local energy resources provided by renewable energy sources. To better manage such an extended system, a new concept for the electrical grid was born: Smart Grid.

Smart Grid, also called smart electrical/power grid, intelligent grid, ingelli-grid, future-grid or inter-grid [19], is considered as a modern electric power grid infrastructure [20]. It is designed to be the grid of the future [21]. Smart Grid is a self-healing distributed electrical network including dynamic optimization techniques that use real-time measurements to detect and optimize electrical losses, increase reliability, and combine different energy sources to fulfill energy needs [22, 23, 19, 20], while minimizing greenhouse gas emissions [24].

Nevertheless, a critical step to enhance Smart Grid energy efficiency is to extend

smart into energy consumer sectors (transportation, buildings, factories, etc.) [25].

## 2.3 Building Efficiency

As mentioned in Section 2.2, energy consumption grows in the world, therefore, many problems appear. On the one hand, the excessive use of non-renewable energy, which means that the required time for the energy sources to be renewed (petroleum for example) needs a very long time (thousands of years), compared to the time of its consumption (days, months, or years). On the other hand, the use of these energy sources causes the greenhouse gas effect and global warming [26], which is very dangerous for humankind.

Among the top energy consuming sectors, buildings remains to be one of the largest share on final energy use and greenhouse gases emissions [27, 28, 29, 30]. In 2018, the building sector represents 36% of the global final energy use, and emitted 39% of  $CO_2$  [30].

Building energy consumption depends on many factors, such as physical characteristics, also known as building characteristics (envelope, maintenance, orientation, etc.), ambient characteristics (space heating and cooling, lighting, appliances, etc.), and the occupant behavior (occupancy, movement, etc.) in the building. In this dissertation, we are focusing on studying occupant behavior and its impact on building an indoor environment and, therefore, energy consumption.

The final energy consumption by use in buildings, in 2018 are [31]:

- Space heating/cooling: They represent the greatest sources of energy consumption in buildings [32, 33, 27, 34]. Both count for 45.16% of global building energy consumption. Studying HVAC (for Heating, Ventilation, and Air-Conditioning) energy optimization is very important, not only because it represents the highest energy consumers within the buildings, but because it directly influences the building occupant's thermal comfort, and it takes time to correct any discomfort.



For example, if an office occupant thermal comfort is fixed to be  $22C$  and if the occupant gets into his/her office room and find the temperature is  $18C$  the HVAC system needs time (several minutes or several tens of minutes) to heat the office room to reach the required temperature. During this time, the occupant will feel this thermal discomfort. Many studies aim to predict HVAC system energy consumption to optimize energy consumption and preserve occupant's thermal comfort [35, 36, 37, 38, 39, 40, 41, 42]. This shows that there is a real interest in studying and reducing space heating and cooling.

- Water heating: Near 16.62% of building energy consumption is for water heating. Many research studies [43, 44, 45, 46, 46] propose to use renewable energy (Photovoltaic, biomass, wind turbine) to heat water and store it in a water tank to reuse it when necessary. They mainly focus on replacing non-renewable energy with renewable energy to decrease the greenhouse gas impact and energy cost. However, there is not an energy consumption reduction here.
- Lighting: 8.46% of the building energy consumption is used by light. The use of the latter can be impacted by many factors, such as the building envelope design [47], or occupant behavior [48, 49]. Therefore, some studies propose to improve the building envelope, or understand occupant's behavior, to minimize the use of light. Furthermore, some research studies propose to use renewable energy sources, such as photovoltaic systems, instead of non-renewable energy [50], or using natural light instead of artificial light while saving visual comfort [51]. However, as lighting can instantly be turned-on, the occupant's lighting discomfort duration can be reduced by, for example, using a Passive Infrared occupancy sensor to turn-on the light if an occupant enters a room. This is one of the reasons that fewer research studies are dealing with the building's lighting optimization problem, compared to those studying building occupant's thermal comfort problems. The latter can last longer, and rapidly leads to a real problem with the buildings' occupants.
- Cooking, appliances, and other electrical equipment: The energy used for cook-

ing, and appliances and other electrical equipment in the building are estimated to be 15.38% and 14.38%, respectively. Some research work-study the occupant use of appliances and its impact on building's energy consumption [52], or replacing conventional electrical appliances with energy-efficient appliances [53, 54, 55]. Generally, cooking energy consumption requires material solutions. For example, using an induction cook-top instead of electrical ones helps reduce wasted heat. For appliances, nowadays there are smart sockets that help occupants remotely controlling them and there are sockets able to detect human activity on appliances, and turn them off when there is no activity detected.

## 2.4 Smart Building VS Intelligent Building

In this section, we discuss the importance of data in Buildings in subsection 2.4.1, introduce Smart Buildings in subsection 2.4.2 then, explain what is BEMS (for Building Energy Management System) in subsection 2.4.3.

### 2.4.1 From Building Data Collection to Intelligent Building

To optimize building energy consumption, we need to observe, understand, and detect energy consumption flaws, by collecting data and information within the building then, propose actions or recommendations. Primary data collection represents an important element in many research projects [56].

Many data and information collection methods exist: telephone interviews, questionnaires and surveys, face-to-face meetings, focus groups, etc. [57, 58, 56, 59], have been used. Ensuring that gathering data is realized in a scientific and standardized manner helps to obtain high-quality research results and thus, credible findings [58]. For example, face to face methods may be ineffective, especially if the interviewee has forgotten the answer to some questions about old details, or hide information because he/she is afraid of prejudice. This can lead to having missing or erroneous data.

To enhance the validity, accuracy, reliability, and credibility of research results, data collection methods need to be improved [56]. Thanks to the Internet of Things (IoT), building's data gathering has become much easier, and more accurate, and flexible [60, 61]. IoT allows automating data collecting processes using sensors within buildings and their surrounding areas. Sensors are widely used in buildings to collect indoor data, such as room temperature, air quality, occupancy state, and outdoor data, such as temperature, wind speed, humidity, etc. Using sensors are programmable and adaptable. They help collect accurate data without the need for human interaction. The collected data can be automatically sent to a control center, where they can be processed and used for controlling room temperature, air quality, security intrusion detection, etc. Furthermore, other technologies, such as Big Data, Cloud Computing, Edge Computing are often used to help store and process data collected by the sensors.

In addition to sensors, another type of device can be deployed to help use the data collected by the sensors and send them to the control center, then, in an automated manner, control building's devices, such as heaters, coolers, lights, appliances, etc. These devices are known as actuators. It operates in the reverse direction compared to sensors. It takes orders from the control center and operates on devices, to turn-on a heater, open a window, turn-off the light, etc. Figure 2.1 summarizes the interaction flow from sensor data collection to actuator triggering within the devices. A building deploying this kind of system is called Intelligent Building (IB) [62, 63, 64]. The building can collect and analyze data automatically and then, interact with the building's devices (heater, light, etc.).

### **2.4.2 Smart Building**

A Smart Building is a responsive, adaptable, and flexible building [65]. Unlike an Intelligent Building, in a Smart Building, the occupants can interact with the system and be part of the decision making process, using user interfaces with the devices to, for example, track their energy consumption, consult and change rooms ambient temperature,

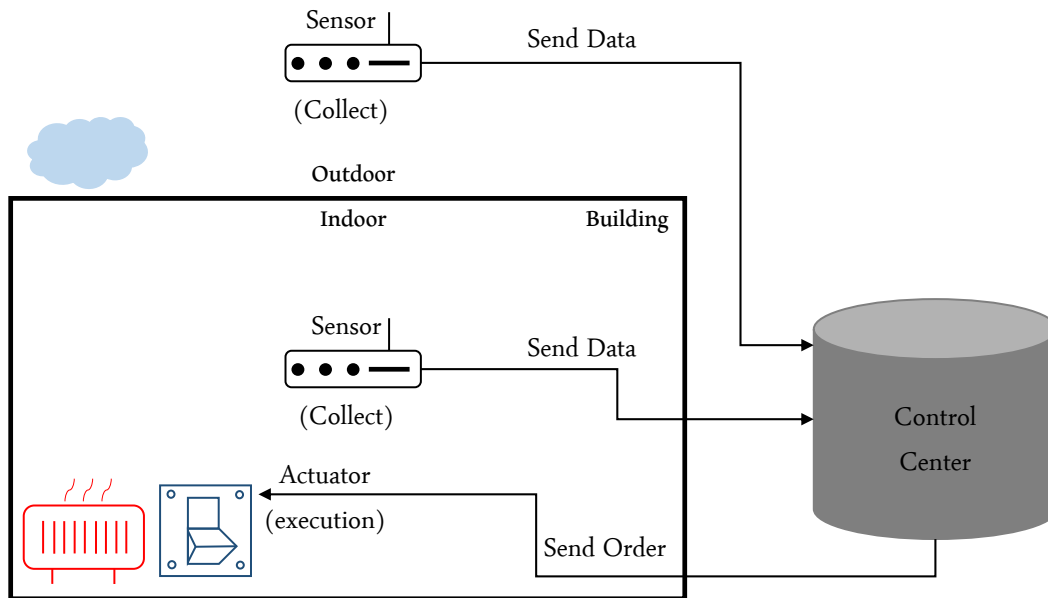


Figure 2.1: From Sensor to Actuator.

monitor air quality or many other factors, etc. Furthermore, in a Smart Building, the control center can deploy Artificial Intelligence algorithms to use data to understand, learn, and predict future decisions. In another word, Smart Building makes the human in the heart of the system. It becomes a human-centric system, in which the occupant can interact with most of the building's devices using smart objects such as smartphones, smartwatches, etc. when He/She is within the building or even far from it.

### 2.4.3 Building Energy Management System

All the previous improvements brought to buildings, cited in the previous, show the importance to have a structured management system that runs the full process to control the building, from the data collection to decision making and occupant assistance. This system is called Building Energy Management System (BEMS). It is composed of a combination of strategies and methods used to improve building performance, efficiency, and energy utilization and optimization [66]. BEMS allows key energy management tasks implementation, such as supervising energy cost, detecting energy-wasting sources,

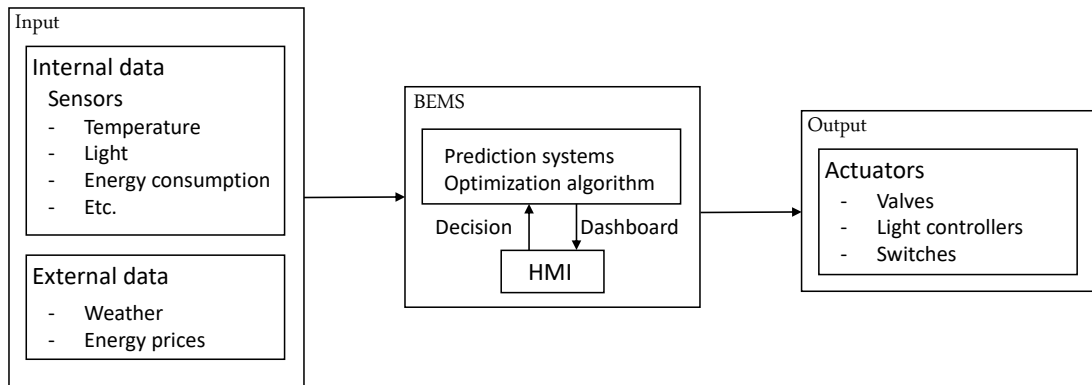


Figure 2.2: BEMS: from data to decision.

automating demand and response approaches, study occupant comfort aspects, and clarifying energy consumption information [67].

In BEMS, occupant interaction is important thus, we can find dashboards using Human Machine Interfaces (HMI) deployed in the system to allow occupants to interact with it. Figure 2.2 summarizes the BEMS interaction with building devices and the occupants. One missing point in current BEMS to allow the full exploitation of the Smart Building concept is that there does not yet allow the integrating of automatic control loops based on Artificial Intelligence algorithms and allowing to implement future decisions from collected data.

## 2.5 Building occupant behavior

We cannot talk about building energy optimization without mentioning the occupant. As shown in section 2.3, the main building energy consumers depends on the occupant's behavior, such as space heating and cooling, and lighting, whose counts for 45.16% and 16.62% of global building energy consumption, respectively. Building heating and cooling aim to satisfy occupant's thermal comfort, and lighting ensure the occupant's visual comfort. Therefore, while optimizing building energy consumption, we have to take into consideration the occupant comfort and, because its behavior influences building energy

consumption in various manner [68, 69, 70, 71], we need to closely understand this behavior to not decrease his/her comfort, otherwise, we risk having strong opposition to changes from him/her. Moreover, occupant behavior is an important factor for building design and retrofit technologies evaluation [72, 71]. So, its understanding can also serve that purpose. As such, it is important to understand, learn, and predict occupant behavior to minimize building energy consumption [73, 74, 75].

Besides, Occupant behavior represents an enormous source of uncertainty in building energy modeling [76]. The occupant behavior parameters that can be taken into account for optimizing building energy consumption are presence, movement, window operation, shading operation, thermostat adjustment, lighting operation, appliance use, clothing adjustment, etc. [75].

## 2.6 Thesis positioning

Based on the discussion presented earlier in this chapter, we need to have a trade-off between two main factors. Building energy optimization and conserving occupant comfort. The state-of-the-art review shows that there is a need to develop more efficient and robust solutions to achieve these two goals. This dissertation proposes a set of tools contributing to reaching these two goals. These contributions are summarized in the following:

- The development of a room-based method to recognize the adequate set of data and data collection frequency, within this set allowing to predict accurately the occupant behavior. The method has the following advantages:
  - Predict building's occupant behavior with the least possible amount of data. It is important to use data in occupant behavior understanding for building energy optimization, but it is even more important to just use the adequate data and for it. Therefore, the question "what data do we need?" is a vital question. In fact, after fixing the goal of energy consumption minimization,

the need for data is immediate and that is why it is important to use the adequate data-set and avoid random use of data.

- Minimize building energy consumption by minimizing sensors. Deploying sensors within the building to understand occupant behavior is important and very useful in energy optimization. Nevertheless, installing sensors leads to capital expenditure (CAPEX) and operation expenditure (OPEX). Therefore, minimizing the number of used sensors implies a minimization (CAPEX and OPEX). Furthermore, in some sensors (example: temperature, humidity), data collection frequency is programmable and, indeed, the shorter data collection frequency, the higher energy sensors consume. Consequently, it is important to select the longest data collection frequency possible to keep the sensor last for long without human interactions to change batteries (i.e. OPEX reduction). Nonetheless, it is vital not to select a random long frequency just to minimize energy consumption, but, we need to select the data collection frequency that allows for accurate occupant behavior prediction. As a result, we need to find a trade-off between the number of sensors to deploy and the data collection frequency.
- A higher abstraction level for a better occupant behavior understanding. Occupant behavior within the building is a combination of its behavior within all building rooms. Therefore, we propose a higher abstraction level representation of the building to better visualize the inter-relations between occupant behavior and all building rooms. We propose to represent the building as a graph, where rooms are represented as the graph's nodes and the occupants' possible movements as the graph's edges. Figure 2.3 illustrates this abstraction level. The presented abstraction has the following advantages:
  - The relation between occupant movement behavior and room occupancy. Generally, room occupancy is studied for each room, separately. But, in reality, the occupant moves within and among rooms, and if the occupant leaves a

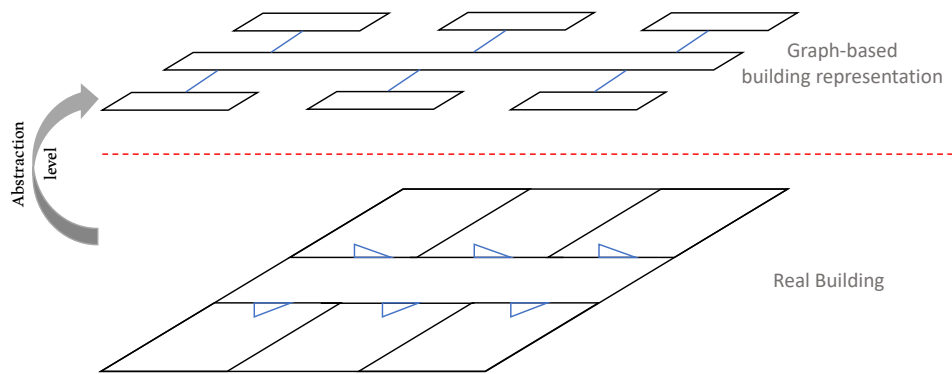


Figure 2.3: Graph-based building representation.

- room (unoccupied room), he/she will go into another room (occupied room). Therefore, room occupancy prediction can be studied based on the relation between building rooms, and the prediction that one room will be occupied may imply that another room will be unoccupied.
- Using graph-based representation allows applying graph-based algorithms such as graph-mining or graph theory toolset.
  - Generic representation of the building. Adding to what has been said, this graph representation can be used for other possible purposes. For instance, it allows different representations of all building rooms such as including virtual separation of rooms. For example, in a hallway, there are two HVAC systems, we can represent this hallway as two separate rooms to study the movement frequency of each of them and, perhaps, propose to physically separate the hallway into two different hallways if needed.
  - The development of an occupant-based energy and comfort adaptation method. As described in Section 2.5, occupant behavior has a direct and indirect impact on energy consumption within buildings. Nonetheless, minimizing energy can have a negative impact on the comfort of the occupant. Therefore, in our method, we propose to use a graph-based occupant behavior prediction method to optimize the building's energy and occupant's comfort. The main advantages of the proposed



method are as follows:

- Room-based occupant behavior prediction for building energy optimization seems to be an efficient solution. Indeed, if the occupant stays in his/her room for a long duration (his/her own office for example), predicting his/her behavior, based on room information only, could be efficient. Nevertheless, if the occupant often changes his/her room, room-prediction risks being less accurate. Therefore, the occupant-centered prediction method helps learn the behavior of each occupant within the building and then gives more accurate prediction results.
  - By understanding the behavior of each occupant on his/her individual and shared rooms, optimization decisions can be adapted based on new behavior changes.
- Improve the building's occupants' comfort by including real-time data. Occupant comfort (especially thermal comfort) is an important factor that needs to be preserved while minimizing the building's energy consumption. Unfortunately, prediction errors can lead to discomfort periods if they last longer. Therefore, we propose to use associate real-time data information to correct prediction errors.
    - Real-time error detection for an imminent correction. While minimizing the building's energy consumption, indoor comfort risks being impacted. Occupant's behavior prediction decides the state of some building's equipment such as HVAC state (ON/OFF). If the decision during distant time intervals, any wrong decision (especially false-negative decision) can impact the occupant's comfort. For example, if the occupant is inside a room that has been predicted as non occupied, the HVAC system will be turned-off and, unfortunately, the occupant will have thermal discomfort until the next decision time stamp. Therefore, we propose to, in addition to prediction-based decision making, use real-time data to detect and correct any wrong decision, thus increase the occupant's comfort.

## 2.7 Conclusion

In this chapter, we reviewed the problem of increasing energy consumption then, we have defined the main energy consumer sector then, the main reason for energy consumption in the building. Selecting the adequate set of sensory data and the associate data collection frequency is important for optimizing building energy consumption as well as sensor deployment and maintenance cost. Furthermore, understanding occupant behavior and the relation between building rooms helps predict occupant's behavior and thus optimizing energy while conserving occupant's comfort. Answering these two objectives is the target of this dissertation. The different contributions will be details in the next three chapters.



## Chapter 3

# Understanding Building Occupants’ Behavior Based On Artificial Intelligence and IoT

### Contents

---

<b>3.1</b>	<b>Introduction</b> . . . . .	<b>44</b>
<b>3.2</b>	<b>Context and Motivations</b> . . . . .	<b>45</b>
<b>3.3</b>	<b>Related Work</b> . . . . .	<b>46</b>
<b>3.4</b>	<b>Data Reduction Model Description</b> . . . . .	<b>48</b>
3.4.1	Feature Selection-based Data Collection Period Selection Algorithm (FS-DCP) . . . . .	49
<b>3.5</b>	<b>Experiments</b> . . . . .	<b>51</b>
3.5.1	Dataset description . . . . .	51
3.5.2	Algorithm parameters ( $P_i, F_j, APA, MLC$ ) . . . . .	52
3.5.3	Results and discussion . . . . .	54
<b>3.6</b>	<b>Summary of the Results</b> . . . . .	<b>59</b>
<b>3.7</b>	<b>Conclusion</b> . . . . .	<b>60</b>

---

## 3.1 Introduction

Building energy consumption depends on many factors, such as occupant behavior and occupancy. Many works studied building occupancy modeling and its impact on energy consumption. Building occupancy prediction requires understanding building information, extracted from datasets created from building collected data. These datasets can be used to identify, train a model, and predict occupancy. Besides, instead of using surveys and questionnaires, the IoT market has been developing sensors, to be deployed within buildings and their surrounding areas, to help collect building data automatically.

In this context, existing datasets have been empirically built without considering the relevant sensor types and the data collection frequency for building occupancy modeling. The random deployment of sensors is cost and money consumer, and data collection frequency impacts sensors' energy consumption and batteries life. Therefore, in this chapter, we aim to solve the data collection period problem and to optimize the sensor (i.e. data type) selection process, with the lowest complexity. We introduce an approach to select the data collection period and the relevant and most dominant sensors for building occupancy prediction with satisfying accuracy. Our approach uses Feature Selection and machine learning classifier algorithms, which are applied to different data collection periods, starting from 1 minute to 60 minutes.

For the experiment, we use a real dataset from one-year data collected from many rooms in E.ON Energy Research Center in Aachen Germany. We compare 5 different machine learning classifiers (Random Forest Classifier, Decision Tree Classifier, Extra-Trees Classifier, Gaussian Naive Bays, and Multi-Layer Perceptron Classifier). Results show that it is possible to build an occupancy predictive model with Random Forest having an accuracy of at least 90%, by using 8 sensors collecting data at a 20-min interval, or 5 sensors collecting data at a 15-min interval.

The remainder of the chapter is organized as follows. We first present the context and motivations in Section 3.2. In Section 3.3, we discuss related work. We detail our

proposed model for data selection and data collection period optimization in Section 3.4. Section 3.5 covers experiments and obtained results, and we discuss our contribution assessment in Section 3.6. Section 3.7 concludes the chapter.

## 3.2 Context and Motivations

Theoretically, the Internet of Things and Artificial Intelligence might help in understanding building occupants' behaviors. But, in practice, there are many existing sensor types and configuration options. In this chapter, we are focusing on selecting the most relevant sensor types and data collection periods to achieve an accurate and cost-effective occupants behavior prediction. Also, with all available Artificial Intelligence techniques and algorithms, it is important to study, for each context, which is the most appropriate algorithm to use. This is also part of our objective to identify the best performing tool for occupant behavior prediction. To sum up, in this chapter, we are interested in:

- Defining relevant sensor types: Selecting appropriate data types, within all available data collectors (sensors) on the market, is a difficult and crucial task. In fact, the more we deploy sensors within a building, the more it costs in terms of investment and maintenance. Furthermore, and depending on the context and goal, some sensor types are more relevant than others. Since selecting relevant data types within  $n$  sensors is an NP-complete problem, we aim at providing in this chapter a simple and efficient method to select the appropriate data types i.e. sensor set for occupants' behavior prediction.
- Defining the data collection period: Sensors are configured to send data each  $P$  period. This period can vary, from a few seconds to many hours or days. Data gathering time directly influences the sensor's battery lifetime, for wireless sensors, or building energy consumption, for wired sensors. Therefore, it is important, for building maintenance minimization, to reduce this cost by reducing the data collecting period, without negatively affecting the trained model quality. Further-

more, identifying the key parameters to measure is important to avoid unnecessary and redundant data collection, while redundant sensor deployment will cause unnecessary cost and disturbance on occupant behavior modeling [77]. As stated in [77], we need not only to select the relevant sensors but to identify their placement (indoor, outdoor) as well.

With all this in mind, in this chapter, we focus on defining a data reduction mode, then an algorithm for extracting, from the collected set of data, the optimal data features, and data collection frequency for human indoor occupancy prediction. our objective is to build an accurate predictive model for occupancy prediction. To deal with our issues of selecting an optimal set of features, we harnessed tools and approaches from the field of machine learning. To achieve this use first start by ranking the features from the most to the least important ones, using the feature selection method, then, we proceed with a classification model based on 5 well-chosen Machine Learning Classifiers (Random Forest, Decision Tree Classifier, Extra-Trees, Gaussian Naive Bayes, and Multi-layer Perceptron).

To assess our approach, we use a dataset about a building, organized as 10 multi-person offices, belonging to the E.ON Energy Research Center in RWTH AACHEN University-Germany. Building data are about indoor temperatures and humidity,  $CO_2$  level, VOC, windows and door state, and outdoor temperatures and humidity. All these are obtained from different sensors deployed inside and outside the building. These features are then associated with office occupancy. The sensors are configured to collect and send their data at 1 *min*-interval.

### 3.3 Related Work

Many studies have been carried out in the field of occupancy behavior prediction [78, 79, 80, 81, 82, 83], to name a few. Chen et al. [78] introduced a multi-tool for energy-related occupant behavior simulation in buildings, to simulate and visualize occupancy and occupant movement in offices, then generates occupant schedules for each space, and

each occupant.

Yilmaz et al. [82] have developed an approach to bottom-up stochastic occupant behavior modeling for predicting the use of household electrical appliances in domestic buildings. The study is limited to domestic buildings in order to determine the data which influences energy consumption. Wang et al. [84] made use of K-Nearest-Neighbors to track building occupancy distribution and occupant activities. The data is collected with proximity iBeacons. Other occupancy detection techniques use sensors (humidity,  $CO_2$ , temperature) or cameras [85, 86, 87] to check if a room in a building is occupied or not, at a given timestamp or time interval. Other works proposed occupancy prediction techniques to estimate when the building/room is occupied [88, 80, 89, 90, 91, 92].

In the recent paper of Li and Dong [88, 90], two short-term building occupancy prediction approaches have been developed. The data is collected by using motion sensors and is also filtered to generate time-series data in 15-min, 30-min, and 1-hour intervals for different prediction window. The experiments are more about the forecasting period than about the data collection period.

Adamopoulou et al. [89] proposed a Spatio-temporal historical analysis-based real-time occupancy prediction. The data, in this case, is collected using depth-image cameras in real-time, and acoustic and motion sensors. Their objective is to compute occupancy prediction, based on the collected data.

Arief-Ang et al. [93] proposed a semi-supervised occupancy counting based on  $CO_2$  sensor only. In our case, we deal with 9 sensors, including  $CO_2$  sensor, to build a predictive model for building occupancy. In another paper of the same team [94], a method to calculate the number of occupants with different classifiers, to identify sensors with a strong correlation. This work is close to ours, but in their case, they limited the number of features to the top 3 dominant ones, and the time has been divided into segments of 10 minutes, and part of day segments (Morning, Afternoon, Evening and Night). In our case, the time and number of features are the main objectives and outputs of our method, such that the time is measured with the data collection period,



and features are determined based on the best accuracy of the predictive model, obtained by those features. To the best of our knowledge, there is no work which deals with dataset creation in terms of data features (sensors to be deployed), and the data collection period (sensors configuration) in order to consider the cost of sensors maintenance and deployment within a building for occupancy modeling and prediction.

### 3.4 Data Reduction Model Description

Our objective is to reduce the amount of data while computing an accurate enough predictive model. There are two ways to achieve data reduction in our IoT context, namely:

- Reduce data dimensionality by applying feature selection methods,
- Decrease the frequency of the data collection: having the sensors sending their data each minute, then it would be interesting to study the effect of a larger period of data collection on the occupancy of the predictive model we can obtain from the reduced data.

In our case, we combine the two aforementioned ways for data optimization. It amounts to solve the following problems, depending on the priority to attach to data dimensionality and data frequency.

#### Priority to frequency:

The problem can be expressed as follows: “to compute the largest data collection period  $\Pi$ , that provides a predictive model with the smallest set of features, ensuring an Acceptable Prediction Accuracy (denoted by  $APA$ )”. More formally, it can be expressed by the following Formula (3.1):

$$\Pi = \max\{P_i \in \mathbb{N} : \min|F_j| \{F_j : (A_j(M_{jP_i}(F_j)) \geq APA)\}\} \quad (3.1)$$

where:

- $P_i$  is the period of data collection. It is fixed by the user as a multiple of the initial period of collection of the dataset:  $P_i = k * P_0$ .
- $j$  is a function computing the accuracy of the model built upon the set of features  $F_j$ ,
- $M_{jP_i}$  is the  $j$ th model built by a given machine learning algorithm,
- $F_j = F_{j-1} \setminus \min_{imp}(F_{j-1}); j = 1, \dots, |F|$ ; and  $F_0 = F$ , and  $imp$  is the importance of features of  $F_{j-1}$ .

### Priority to data dimensionality:

The problem can be expressed as follows: “to compute the smallest set of data features, that provide a predictive model with a larger period of data collection, having an acceptable prediction accuracy”. More formally, it can be expressed by the following Formula (3.2):

$$\min_{|F_j|} \{F_j : \max\{P_i \in \mathbb{N} : A_j(M_{jP_i}(F_j)) \geq APA\}\} \quad (3.2)$$

Where  $P_i, A_j, M_{jP_i}$  and  $F_j$  have the same meaning as in formula (3.1).

We introduce in the next subsection an algorithm implementing Formula (3.1).

#### 3.4.1 Feature Selection-based Data Collection Period Selection Algorithm (FS-DCP)

In this algorithm, we mainly focus on Data Collection Period, denoted DCP. The goal is to find the largest DCP that guarantees an  $APA$ .

Let us suppose that we have a dataset  $D$  of  $N$  data features, collected each  $P_0$  period. We start by generating data collection periods, denoted  $\{P_0, \dots, P_k\}$ .

For each  $P_i$ , such that  $i = \{0, \dots, k\}$ , we extract from the global dataset  $D$ , the sub-dataset  $D_i$  of data collection period  $P_i$ . Then, we apply the feature selection process on dataset  $D_i$  to rank the features from the most to the least important, according to their degree of relevance. After that, we eliminate the weakest relevant feature, and we generate a model using a Machine Learning Classifier, denoted MLC. We calculate its Model Accuracy, denoted by MA, and we compare it to  $APA$ . We repeat the process, starting from eliminating the weakest feature from the new resulting set of features until we process all the features.

At the end of the algorithm, we obtain the largest period of  $P$  with the corresponding shortest set of features ( $F_P$ ), and the accuracy of the current best model obtained. FS-DCP algorithm is detailed in Algorithm 2.

---

**Algorithm 1** FS-DCP computing algorithm.

---

**Require:**  $\{P_0, P_1, \dots, P_k, F_0\}, D, APA$   
**Ensure:**  $(Max(P), Min(F_P), MA)$   
**BEGIN**  
 $Max(P) \leftarrow P_0$  // Max P Initialisation  
 $Min(F_P) \leftarrow F_0$  // Min F Initialisation  
 $MA \leftarrow 0$   
**for**  $i = 0$  to  $k$  **do**  
     $D_{P_i} \leftarrow$  Generate dataset for  $P_i$   
     $RF_{P_i} \leftarrow Rank(F_{P_i})$  // using Feature selection  
    **for**  $j = 0$  to  $|F_0|$  **do**  
         $F_j \leftarrow F_j \setminus F_{(|F_0|-j)}$  //  $F_j$  are ranked  
         $CM(D_{P_i}, F_j)$  // Generate model  
         $MA(D_{P_i}, F_j)$  // Model Accuracy calculation  
        **if**  $(MA(D_{P_i}, F_j) \geq APA)$  or  $(MA < MA(D_{P_i}, F_j))$  **then**  
             $Max(P) \leftarrow P_i$   
             $Min(F_P) \leftarrow F_j$   
             $MA \leftarrow MA(D_{P_i}, F_j)$   
        **end if**  
    **end for**  
**end for**  
**return**  $(Max(P), Min(F_P), MA)$  **END.**

---

The complexity of the algorithm depends on the complexity of the model generated by a given machine learning classifier algorithm, denoted  $C$ , in addition to the complexity

of the nested loops, such that one is over the list of periods, and the other is over the list of features. Therefore, time overhead has a complexity of  $O(C * n^2)$ .

The algorithm is terminal, which means that each of its executions reaches an end, because the list of periods  $P$ , and features  $F$  are bounded. Besides, it returns a solution in both following cases:

- Case 1: There exists a model with an accuracy  $MA \geq APA$ . In this case, the algorithm returns the greatest period  $P$  with its corresponding smallest set of features  $F_p$ .
- Case 2: There is no predictive model having the accuracy  $MA \geq APA$ , the algorithm returns the greatest period  $P$  with its corresponding set of features  $F_P$  that build a model having the closest accuracy to  $APA$ .

## 3.5 Experiments

We detail in this section the experiments we carried out, and we discuss the obtained results. In Subsection 3.5.1, we describe the dataset. In Subsection 3.5.2 we introduce the algorithm settings, and in Subsection 3.5.3 we detail our results with a discussion.

### 3.5.1 Dataset description

In this experiment, we use a dataset containing data collected for 9 offices located at E.ON Energy Research Center in Aachen, Germany.

The data is collected by using 8 sensors deployed inside and outside the building. The sensors are as follows:

- Indoor data:  $CO_2$ , humidity, temperature, air quality (volatile organic compounds: VOC), door state (open, closed), window state (open, half-open, closed),
- Outdoor: Humidity and temperature.

Table 3.1: Features description.

Feature	Data Type	Measure Unit
$CO_2$	Numeric	parts-per-million ( <i>ppm</i> )
Indoor humidity	Numeric	%
Indoor temperature	Numeric	Degree Celsius ( $^{\circ}C$ )
VOC	Numeric	<i>ppm</i>
Door state	Binary {0, 1}	
Window state	Triple {0, 0.5, 1}	
Outdoor humidity	Numeric	%
Outdoor temperature	Numeric	Degree Celsius ( $^{\circ}C$ )
Week day	1: Mon to 7: Sun	

The aforementioned sensors define 8 attributes or data features; we have added the weekday attribute to refer to the time. The 9 attributes are described in Table 3.1. Data is collected between February and July 2018. As the sensors send their data at different intervals, then a process of data collection period is carried out to unify the data collection interval to 1 minute. After we carried out a data cleaning and preparation process, we obtain a dataset of 70500 samples.

### 3.5.2 Algorithm parameters ( $P_i, F_j, APA, MLC$ )

We instantiated Algorithm 2 with the following parameters:

#### Frequency of collection ( $P$ ):

The set of frequencies we have experimented is  $P = \{1, 5, 10, 15, 20, 30, 60\}$ , to study the influence of the data variations over the time (from the high to the low) on the sensitivity of the predictive model.

#### Data features ( $F$ ):

We consider the set of features described in Table 3.1.

**Acceptable Prediction Accuracy (APA):**

We define accuracy as the percentage of correct occupancy predictions over all predictions obtained from test dataset (1/3 of global dataset). From the user point of view, the higher accuracy, the better. We can consider an APA of at least 90% as acceptable prediction quality of generated model.

**Machine Learning Classifier (MLC):**

As the label of the model is binary, then classification is the suitable machine learning method to use in our case. We have instantiated our method with the following 5 classifiers:

- Random Forest Classifier, denoted by RF,
- Decision Tree Classifier, denoted by DT,
- Extra-Trees Classifier, denoted by ET,
- Gaussian Naive Bayes, denoted by GNB,
- Multi-layer Perceptron Classifier, denoted by MLPC.

We have compared them according to the accuracy of the occupancy predictive model they provide. It is noteworthy that we can use any other classifier algorithm.

**Implementation:**

We have implemented our algorithm with Python programming language, using Scikit-Learn machine learning library.

Table 3.2: *DCP, NF, MA* for each *MLC* algorithm.

MLC	DCP (min)	NF	MA(%)
<i>RF</i>	20	9	90.30
DT	10	9	91.94
ET	15	5	90.16
GNB	20	8	79.34
MLPC	15	9	81.39

### 3.5.3 Results and discussion

We applied algorithm 2 on the cleaned dataset described above for the aforementioned *MLC* with each data collection period  $P_i \in \{1, 5, 10, 15, 20, 30, 60\}$ .

Figure 3.1 summarizes the maximum of the model accuracy obtained for all data collection periods considered in our experiment and for each *MLC* algorithm. We notice that *RF*, *DT* and *ET* classifiers ensure the required *APA* (90%). *GNB* and *MLPC* classifiers do not provide any model with the required *APA*. Their best accuracies are 79.34% and 81.39%, respectively.

Figure 3.2 details the minimum number of features for each *DCP* and *MLC* that ensures the required *APA* or the maximum possible accuracy for the occupancy predictive model.

From Fig. 3.1 and Fig. 3.2, we obtain the sufficient *MA* and the corresponding *DCP* and *NF* for each *MLC*. We sum up the results in Table 3.2, where we notice that *RF*, *DT* and *ET* algorithms satisfy the *APA*, instead of *GNB* and *MLPC*, which do not. We can conclude that the Random Forest classifier (*RF*) is the most suitable algorithm to use since it gives the higher *DCP* that provides a predictive model with an accuracy greater or equal to *APA*. Therefore, we focus hereinafter on the details of the experiments carried out for that algorithm (*RF*) with the different data collection periods.

3.3a-3.3g display the model accuracy variation according to number of features considered for each data collection period  $P_i \in \{1, 5, 10, 15, 20, 30, 60\}$ , respectively.

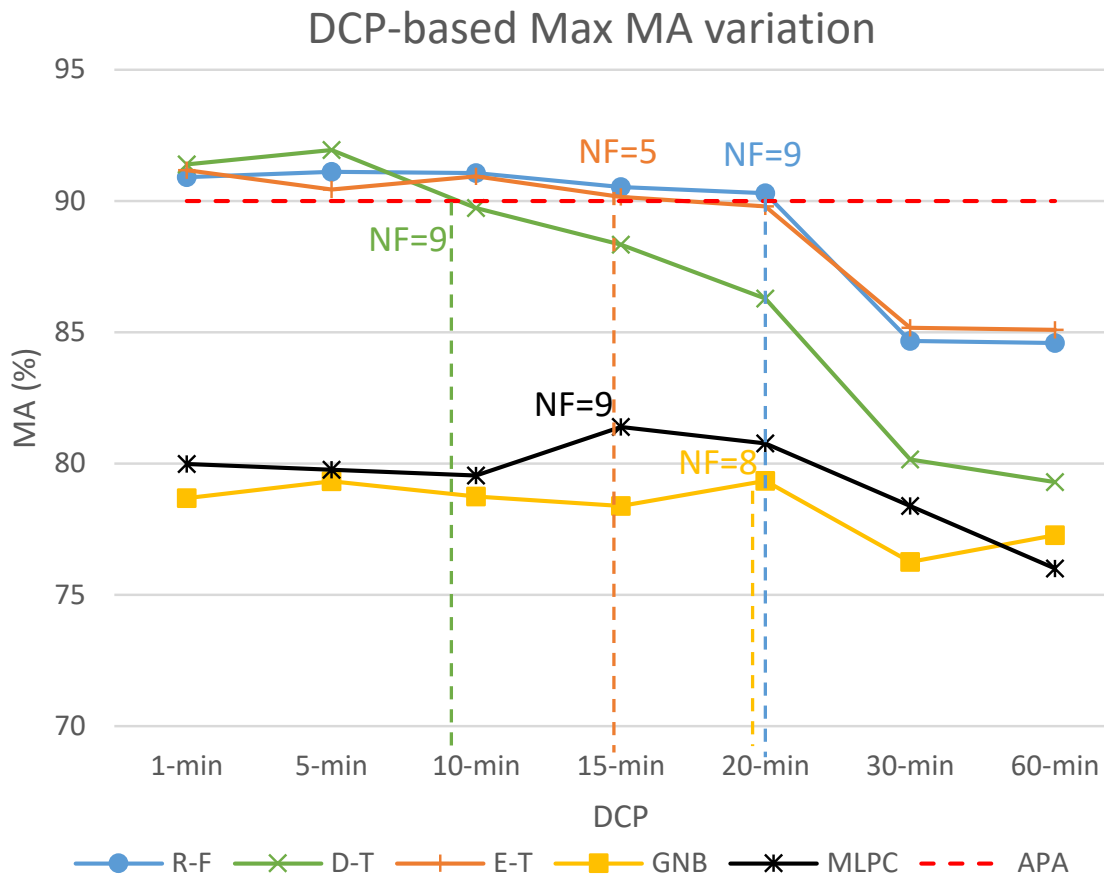


Figure 3.1: The maximum of the model accuracy according to each data collection period for different ML algorithms.



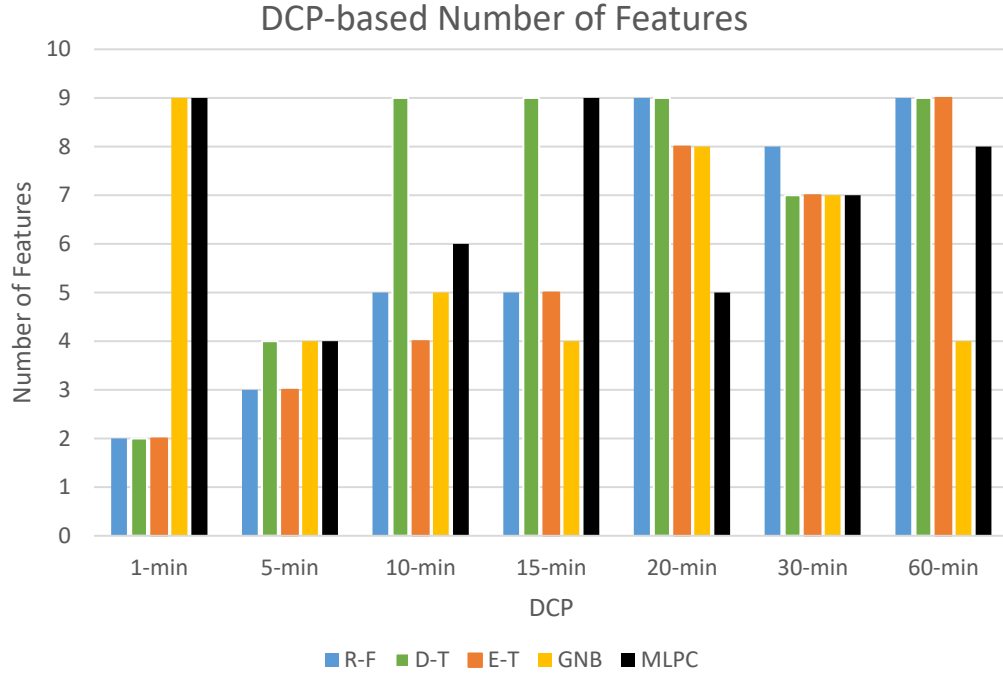


Figure 3.2: Minimum number of features for *DCP* corresponding to the optimal MA (Fig. 3.1), for each ML algorithm.

We noticed that for  $P = 1 \text{ min}$ , 2 features are sufficient to provide a model with an accuracy of 90.9%. 1 minute is the lowest  $P$  and it requires the least number of features comparing to other  $P$ , whereas, for  $P = 20 \text{ min}$  we need to use all the 9 features to obtain a predictive model with an accuracy of 90.3%. For  $P$  greater than or equal to 30 *min*, there is no predictive model achieving the required *APA*. In this case, the algorithm returns the number of features ensuring the closest model accuracy to the required *APA*. For example, for  $P = 30 \text{ min}$ , the number of features is 8, and the accuracy of its corresponding model is 84.67%.

To conclude, the larger  $P$  is, the more features the model requires, until reaching a  $P$  beyond which we cannot build any predictive model (according to the chosen classifier algorithm) satisfying the *APA*. Fig. 3.4 summarizes the maximum accuracy of the model obtained for all data collection periods considered in our experiment. The best trade-off between the model accuracy and the data collection period is provided by  $P = 20 \text{ min}$ . For this period, Fig. 3.3e shows that we need to use all the 9 features.

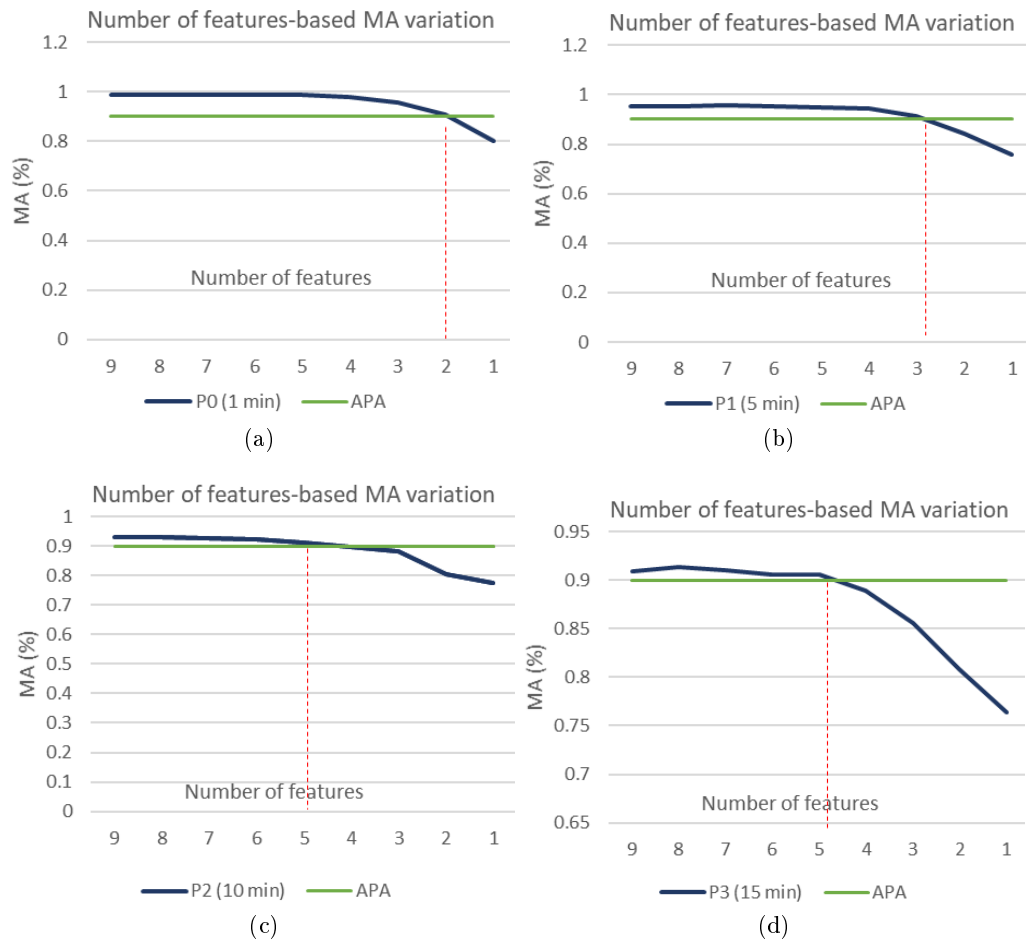


Figure 3.3: The variation of the model accuracy according to the number of features.

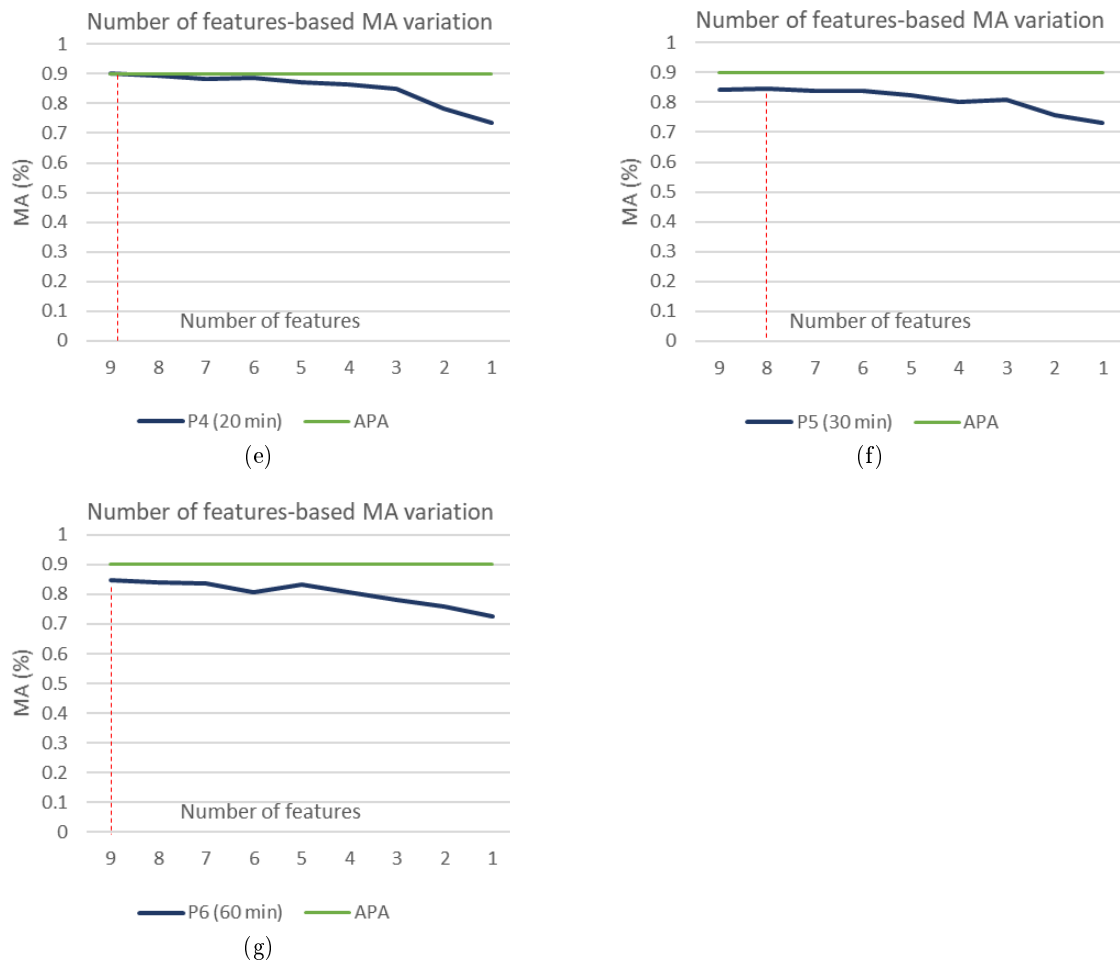


Figure 3.3: (Cont.) The variation of the model accuracy according to the number of features.

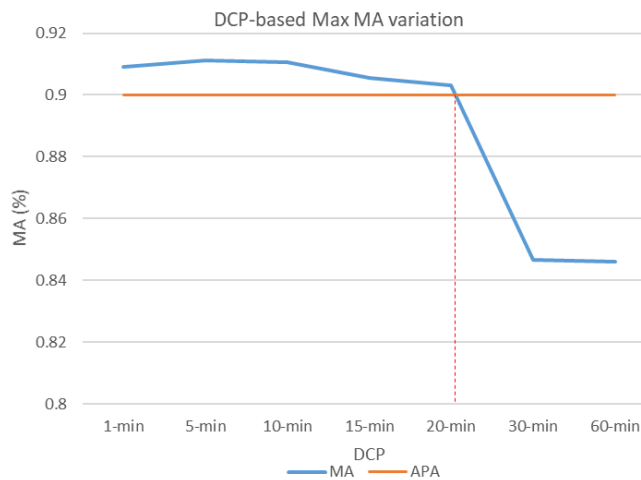


Figure 3.4: The maximum of the model accuracy according to each data collection periods.

**Remark.** We notice in Fig. 3.3d that  $P = 15 \text{ min}$  can be seen as a good deal between the number of features (number of sensors to deploy), which is 5 in this case ( $CO_2$ , indoor humidity, outdoor humidity, outdoor temperature and VOC), instead of 9 for  $P = 20 \text{ min}$ , and the amount of data we need to collect to build a good predictive model (having 90.53% of accuracy). Therefore, it is possible to improve our solution, to consider this kind of trade-off, where the number of sensors is also important to reduce the cost of deployment and maintenance of sensors within buildings.

90%, by using 8 sensors collecting data at a 20-min interval, or 5 sensors collecting data at a 15-min interval.

### 3.6 Summary of the Results

All the above simulation results illustrate the importance of data selection to build a solid building occupants' behavior prediction model. The three main take away messages from our results are:

- The right data type for the right model: minimizing deployed sensors helps min-

imize building energy consumption and sensors deployment costs. Furthermore, selecting relevant data types helps obtain the right information for an optimal prediction model construction. In our experiment, we test the building's occupancy prediction using 8 sensor types which are: Indoor data:  $CO_2$ , humidity, temperature, air quality (volatile organic compounds: VOC), door state (open, closed), window state (open, half-open, closed), and Outdoor data: humidity and temperature. Results show that 5 sensors are sufficient to give an accurate prediction. These sensors are: ( $CO_2$ , indoor humidity, outdoor humidity, outdoor temperature, and VOC). This shows that, in our experiment, more than 44% of the used sensors are not necessary.

- The largest data collection period, the longest battery life: we have proposed a model that helps determine the optimal data collection period to have the longest battery life and optimize sensors' maintenance time and cost while giving an acceptable prediction model. In the dataset we used for our experiment, data were collected each one minute. Our approach shows that collecting data each 15 minutes is sufficient. Thus, the sensors' battery life lasts longer.
- The right algorithm for the right prediction: Many machine learning algorithms can be used for the prediction model construction. Nevertheless, it is important, for each context, to select the most relevant algorithm. In our case, the test results show that Random Forest was the most relevant ML algorithm that provides the most satisfying results, with an optimal number of sensors and data collection period.

### 3.7 Conclusion

In this chapter, we have developed a method to select a data collection period and the relevant sensors to use for accurate building an occupancy prediction. Our approach uses a feature selection algorithm and compares 5 machine learning classifiers (Random

---

Forest, Decision Tree Classifier, Extra-Trees, Gaussian Naive Bayes, and Multi-layer Perception), applied for different data collection periods, starting from 1 minute to 60 minutes from the selected set of sensors. We performed a set of experiments on a real-world data set in 9 offices located at E.ON Energy Research Center in Aachen, Germany. Results showed that Random Forest is the most suitable machine learning algorithm ensuring a good building occupancy modeling built with 9 features (8 sensors and the weekday) associated to a data collection interval of 20 minutes, or with 6 features (5 sensors and the weekday) to collect data at a 15-min interval. These configurations allow training the occupancy predicting model with an accuracy of at least 90%. The method proposed in this chapter is room-centric, i.e. it studies the occupancy behavior of each room, independently from the inter-relation that rooms can have among each other. Indeed, an occupant that leaves a room is more likely to join another room. Such inter-relations are very important in modeling the building occupant's behavior and its understanding in order to achieve energy efficiency and optimized comfort. Proposing such an occupant-centric method, that takes into account the behavior of each occupant within the building as a whole and then, to build a dynamic occupant behavior prediction modeling is the target of the next chapter of this dissertation.



# Chapter 4

## Graph-based Occupant Behavior Modeling in Smart Building

### Contents

---

<b>4.1</b>	<b>Introduction</b>	<b>64</b>
<b>4.2</b>	<b>Context and Motivations</b>	<b>65</b>
<b>4.3</b>	<b>Related Work</b>	<b>67</b>
<b>4.4</b>	<b>Temporal Graph Model for Occupancy Behavior Prediction</b>	<b>68</b>
4.4.1	Building Movement Model	68
4.4.2	Daytime Intervals and Frequent Occupant Movement Behavior Identification	69
	Time interval-based occupants movement timestamp unification	70
	Frequent occupant movement sequence time interval identification	70
	FrEquent MOvement Sequence identification (FEMOS)	72
	Energy optimization process	72
<b>4.5</b>	<b>Experiments</b>	<b>73</b>
4.5.1	Dataset Description	73
4.5.2	Similar occupant movement grouping and daytime identification	74
4.5.3	Frequent Movement Sequence Identification	78



---

4.5.4 Results . . . . .	78
<b>4.6 Summary of the Results . . . . .</b>	<b>79</b>
<b>4.7 Conclusion . . . . .</b>	<b>81</b>

---

## 4.1 Introduction

In the previous chapter, we focused on optimizing occupants' presence prediction by selecting the relevant data types (i.e. sensor set) and the optimal data collection periods. This aims to optimize building energy consumption while keeping the sensor deployment costs and maintenance cost low. The proposed approach is room-centric, which means that it provides room occupancy patterns. Nevertheless, occupancy patterns are related to the occupant movements and their behavior at a room level and, if one of these changes, the room occupancy pattern will no more be accurate. We argue, in this chapter, that a more complete model, considering the building as a whole (i.e. not just one room) would alleviate such an inconvenience.

To alleviate the above-mentioned inconvenience, we introduce, in this chapter, an occupant-centered approach to build a more accurate prediction model. The proposed approach is a temporal graph-based approach centered on occupant movement behavior. It is a higher abstraction level of the building organization and occupant movement behavior representation within the building as a whole (i.e. a set of interconnected rooms and spaces). In this chapter, we propose to:

- Build a higher abstraction of both the building and its usage for a better understanding: In this step, we represent the building as a graph, where building rooms are represented by nodes, and possible direct movements between the rooms are represented by edges. This representation, which uses graph theory, can be used for any building type (habitats, offices, etc.).
- Understand and Learn occupants movements pattern throughout the above-mentioned

representation as well as graph mining algorithms and multi-objective optimization then, build an occupant movement prediction model based on this Spatio-temporal data.

The remainder of this chapter is organized as follows. We start by giving the context and motivations in section 4.2. Section 4.3 sums up some recent works in the domain of building occupant behavior and energy consumption reduction. Section 4.4 details our temporal graph-based building modeling system for building occupancy and energy consumption optimization. Section 4.5 details our experiments to validate our model. A discussion about our obtained results is given in Section 4.6. Section 4.7 concludes the chapter.

## 4.2 Context and Motivations

Building room occupancy status impacts energy consumption, as stated in [83]. Moreover, building rooms may be occupied differently depending on the usage and the importance of the rooms from the occupant standpoint. For example, in an office building, a break room is located at the same floor as offices can be occupied at midday for an hour, but not occupied at all the rest of the day. Therefore, we consider a time relationship linking occupants to rooms, and a location relationship linking rooms to gather, in the sense that some rooms are connected to others, to a hallway, or even to the outside of the building, which influences the occupant's movements and the overall occupancy of the building.

To be able to predict occupancy status for each room at each time interval, there is a need for a model that helps represent both types of information, upon which it would be easier to predict occupancy within accurate time interval division of days, so as an energy optimization process can be performed accordingly. In this context, we can make use of graphs as a powerful modeling tool, because they can better identify occupant movement behavior among rooms, regardless of the type of building (residential, office,

public administration, etc.), and provide a large amount of information to be used for energy optimization.

Therefore, in this chapter, we propose a temporal graph-based approach for occupant behavior modeling and energy consumption optimization, in which nodes represent rooms of a given building, and edges represent occupant movements among the rooms of the building. A timestamp is also attached to edges to carry the time information. Then, a graph mining algorithm is used to extract the most frequent subgraphs and, thus, allows us to learn the best temporal coverage of the rooms in the building. Furthermore, the time information is exploited in a hierarchical clustering algorithm to define the best division of days into significant time intervals, which maximizes the comfort and minimizes the energy loss thanks to a multi-objective optimization problem we derived as well as its resolution.

We carried out an experiment on a synthetic dataset representing 4 weeks of 4 person movements inside an office building at office hours (from 8:00 am to 6:30 pm), organized in 8 rooms equipped with HVAC systems. The dataset split into training and test datasets, such that, the first one contains 3 weeks, and the second contains the fourth week. The obtained results show that our approach can reduce the heating/cooling period by up to 62.21% with an average of 58.16%, comparing to the traditional permanent heating approach, leading to an energy consumption reduction. By doing so, our approach still guarantees an average mean of 71.28% of occupant's comfort. Moreover, we compared our graph-based approach to a machine learning approach for occupant energy consumption prediction, an approach similar to the one proposed in [95] and where an extreme deep learning approach is defined. The obtained results show that our approach performance is of the same order of magnitude as the one based on machine learning, but it is simpler. Also, our approach allows visualizing occupant behavior and provides coverage information involving many rooms at once.

### 4.3 Related Work

Many studies have been carried out in the field of occupancy behavior prediction. The main proposed approaches have been introduced in [78, 81, 79, 80, 82, 83]. Chen et al. [78] introduced a multi-tool for energy-related occupant behavior simulation in buildings, to simulate and visualize occupancy and occupant movement in offices, then generate occupant schedules for each space, and each occupant. Yamaguchi and Shimoda [81] proposed a stochastic discrete-event model to generate occupants' activities at home to be used in community-/urban-scale energy demand models. Sun and Hong [79] used building performance simulation to study the impact of occupant behavior on energy conservation measures, based on behavior styles (austerity, normal, and wasteful) to develop three baseline models which have been used by decision-makers to reduce the risks of energy retrofit associated with the occupants. Wang et al. [84] made use of K-Nearest-Neighbors to track building occupancy distribution and occupant activities.

Other occupancy detection techniques use sensors (humidity,  $CO_2$ , temperature) or cameras [86, 87, 85], to check if a room in a building is occupied or not, at a given timestamp or time interval. Other works proposed occupancy prediction techniques to estimate when the building/room is occupied [89, 90, 88, 91, 80, 78]. In the recent paper of Li and Dong [88], two short-term commercial building occupancy prediction approaches have been developed. The first is based on a Markov model including a change-point analysis, and the second is a modified random sampling approach with 24-h ahead prediction. Li and Dong [90] proposed a Markov model and a moving learning window for occupancy presence prediction. The method was compared to a modified probability sampling, Artificial Neural Network (ANN), and Support Vector Regression (SVR) methods. Results showed that the Markov model provides up to 15% correctness comparing with the other methods. Massana et al. [91] studied the importance of occupancy indicators, such as binary occupancy, daily profile, and hourly profile on building occupancy prediction, using many building scenarios and data, such as electrical load, temperature, calendar, classroom devices, etc. Adamopoulou et al. [89] proposed

a Spatio-temporal historical analysis-based real-time occupancy prediction. The results show that the narrow period time-based implemented prediction method (15 mn, 30 mn) gives more accurate prediction results than the large time prediction period.

The above-mentioned machine learning approaches to model and predicts building occupant behavior, based on building appliances, windows, or on the impact of occupancy on energy consumption. Although their simulation results are promising in terms of accuracy of prediction, still they are limited in the sense that they consider only one output feature, such as a given office occupancy status prediction, while in our case, we need to combine both time and location to predict the rooms that would be occupied in different time intervals in a day, in order to optimize energy consumption. We also differentiate rooms occupancy duration and room passage frequency, to distinguish among rooms having a low passage frequency but a high occupancy period, which is easier to model in the temporal graph-based approach we detail in the next section.

## 4.4 Temporal Graph Model for Occupancy Behavior Prediction

In this section, we first introduce our temporal graph model for building representation (Subsection 4.4.1). Then, we develop our frequent movement identification process using graph theory and graph mining (Subsection 4.4.2).

### 4.4.1 Building Movement Model

Let us suppose a building with  $V$  rooms and  $O$  occupants. Building movements of an occupant  $O_i$  are seen as a temporal oriented graph denoted by  $G_{O_i} = (V, E_{O_i}, \mathcal{T}_{O_i})$ , such that  $V$  is the set of rooms (nodes),  $E_{O_i}$  is the set of movements of  $O_i$  among rooms (edges), and  $\mathcal{T}_{O_i}$  is a function attaching timestamps of occupant  $O_i$  to the edges  $E_{O_i}$ , as labels. This model is flexible enough to represent physically or logically spaces, such as a long hallway with two heaters can be represented as two rooms  $R1$  and  $R2$  connected

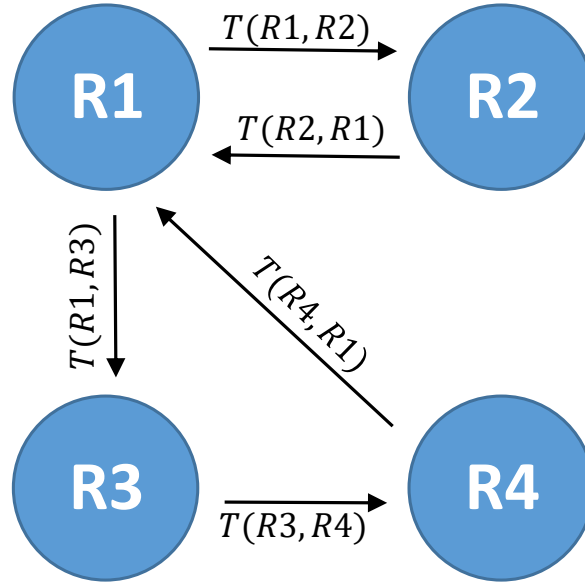


Figure 4.1: Example of building movements of an occupant.

with an edge  $(R1, R2)$ .

**Example.** Fig. 4.1 illustrates a building movement graph  $G_{O_i}$ , where the rooms are represented by the nodes  $R1, R2, R3$  and  $R4$  and the movements are also attached with timestamps:  $\mathcal{T}_{O_i}(R1, R2), \mathcal{T}_{O_i}(R2, R1), \mathcal{T}_{O_i}(R1, R3), \mathcal{T}_{O_i}(R3, R4), \mathcal{T}_{O_i}(R4, R1)$ .

#### 4.4.2 Daytime Intervals and Frequent Occupant Movement Behavior Identification

Our approach aims to find the optimal room occupancy in a day. First, we need to split a day into time intervals, and then, identify which set of rooms the occupants could visit in each time interval. Our approach consists in the following four steps, from data modeling to decision making:

- Time interval occupants movement timestamp unification
- Frequent movement sequence time interval identification
- Frequent movement sequence identification

- Energy optimization process

These four steps are detailed in what follows.

### Time interval-based occupants movement timestamp unification

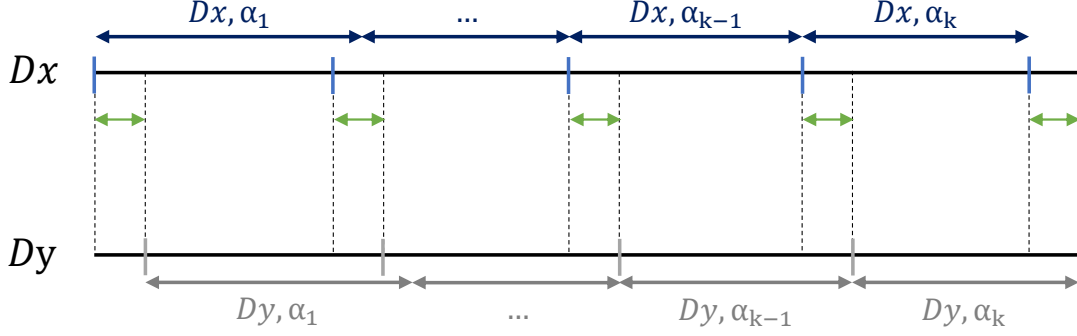
Since graph mining algorithms classify two movements as different, if they do not have the exact label, then we make use of a hierarchical clustering method to group timestamp of occupants' movements, which have similar source and destination rooms, having timestamps close to each other, whatever the day. Thereafter, we set the same timestamp (hour and minute) for each group.

Remark: Since our approach is based on movement behavior among building rooms, the occupant's labels are removed. This helps preserve the occupant's privacy.

**Example.** Let  $M1_{O_1} = ((R1, R2)_{O_1}, [D_1 : 18h00])$  and  $M2_{O_1} = ((R1, R2)_{O_1}, [D_2 : 18h02])$  be two movements for the occupant  $O_1$ . The hierarchical clustering method puts them into the same group. Therefore, they are attached with the same timestamp  $Min(TI_{M1,1,O_1}, TI_{M2,1,O_1})$ , which is  $[18h00]$ . Thus, the two movements become  $M1 = ((R1, R2), [D_1 : 18h00])$  and  $M2 = ((R1, R2), [D_2 : 18h00])$ .

### Frequent occupant movement sequence time interval identification

In this step, we compute the optimal daytime interval decomposition that optimizes frequent movement sequence detection. We start with dividing each day  $D_x(O_i)$  for the occupant  $i$  into time intervals. The  $j^{th}$  time interval of day  $D_x(O_i)$  is denoted by  $TI_{D_x,j(O_i)}$ . Then, we gather occupant movements  $MTI_{D_x,j}$  for each time interval  $TI_{D_x,j}$  into groups. To do not divide randomly days into intervals, which gives many insignificant division possibilities, we use a hierarchical clustering method for each day, which groups occupant movements close to each other, based on time distance. We denote by  $[MinTI_{D_x,j}, MaxTI_{D_x,j}]$  the  $j^{th}$  time grouped interval of day  $D_x$ .

Figure 4.2: Distance between two days with  $k$ -intervals.

Since we have  $d$  days and  $o$  occupants, then we will have  $d * o$  possible time interval groups. Instead of identifying occupant movements for all  $d * o$  groups, we eliminate similar days time intervals by grouping days which have the same number of interval decomposition. Then, we compute for each group of days a representative day  $D_x$ , which has the smallest distance to the center of the group  $d_{D_x} = \text{Min}(d_{\text{group}}) \leq \varepsilon$ , such that  $\varepsilon$  is the maximum distance allowed for a day to be a representative candidate.

Fig. 4.2 illustrates the principle of distance computation between two days divided into  $k$  intervals. Each interval  $i$  of day  $D_x$  (except the first interval  $i = 1$  and the last interval  $i = k$ ) has  $\text{Min}_{D_x,i} = \text{Max}_{D_x,(i-1)}$  and  $\text{Max}_{D_x,i} = \text{Min}_{D_x,(i+1)}$ .

We compute the distance by using the quadratic mean which allows aggregating both positive and negative values. Similarity between two intervals  $\alpha_1, \alpha_2 \in \{1, \dots, k\}$  of day  $D_x$  and  $D_y$  of Fig.4.2 is calculated as follows:

$$\|D_{x,\alpha_1} - D_{y,\alpha_2}\|_2 = \sqrt{\frac{(m_{x,\alpha_1} - m_{y,\alpha_2})^2 + (M_{x,\alpha_1} - M_{y,\alpha_2})^2}{2}} \quad (4.1)$$

To calculate similarity between two days  $D_x$  and  $D_y$ , we generalize Formula (4.1) to  $k$  intervals as follows:

$$\|D_x - D_y\|_2 = \sqrt{\frac{\sum_{i=1}^k ((m_{x,i} - m_{y,i})^2 + (M_{x,i} - M_{y,i})^2)}{2k}} \quad (4.2)$$



### FrEquent MOvement Sequence identification (FEMOS)

We use a graph mining algorithm such as gSpan to identify, for each time interval computed in the previous step, its frequent occupant movement sequences. gSpan as any Graph Mining algorithm requires minimum frequency ( $MF$ ) information to extract all *FEMOSs* that appear at least in  $MF\%$  graphs of the dataset, considered as frequent. Then, we eliminate from graph mining algorithm result all frequent sequences  $FS_x$  which are sub-sequences of other sequences and their frequencies  $F_x$  is less than or equal to the highest sequence frequency. This helps converge the selection decision. The elimination rule is expressed as follows:

$$Eliminate(FS_x) \Leftrightarrow (FS_x \subset FS_y) \wedge (F_x \leq F_y) \quad (4.3)$$

### Energy optimization process

In this last step, we solve the decision optimization problem of turning ON/OFF appliances such as HVAC system in rooms, while optimizing building energy consumption and occupant comfort. The optimization problem is formulated as follows:

$$\begin{aligned} & Min(W_E + L_C) \quad (4.4) \\ s.t. & \left\{ \begin{array}{l} MinHT \leq TI_{Dx,i} < MaxTI \\ 0 \leq N_{UHRDx,i} \leq MaxRN \\ 0 \leq V_{URDDx,i} \leq MaxTI \end{array} \right. \end{aligned}$$

such that  $W_E$  represents wasted energy, and  $L_C$  lack of occupant comfort.  $W_E$  and  $L_C$  are calculated as follows:

$$W_E = \sum_i (N_{URDx,i} * TI_i) \quad (4.5)$$

$$L_C = \sum_i V_{URD_i} \quad (4.6)$$

with  $N_{UHRD_{x,i}}$  represents the number of unoccupied but heated/cooled rooms in day  $D_x$ , and time interval  $TI_{D_x,i}$ ,  $V_{URDD_{x,i}}$  represents the duration of occupied but unheated/uncooled rooms in day  $x$  and time interval  $TI_{D_{xi}}$ , and  $TI_{D_{xi}}$  is the time interval duration  $i$  in day  $x$ .

If two-time intervals tie (equal minimal), then we compare their duration of energy consumption, denoted  $D_{EC}$ , computed as follows:

$$D_{EC} = \sum_i (V_{HRND_{xi}} * TI_i) \quad (4.7)$$

where  $N_{VHRND_{x,i}}$  is the number of visited and heated rooms.

## 4.5 Experiments

For our experiments, we considered the movements of 4 occupants in an office building composed of 8 rooms (Fig. 4.3). Fig. 4.3a illustrates the building structured as rooms and doors. We extract its graph representation in Fig. 4.3b, such that rooms become nodes and doors become edges allowing possible movements.

In Subsection 4.5.1, we describe our dataset. In Subsection 4.5.2, we detail the process of similar occupants movement unification and daytime identification. In Subsection 4.5.3, we compute the process frequent movement sequence identification. We discuss the results obtained in Subsection 4.5.4.

### 4.5.1 Dataset Description

We have generated a 4-week dataset of four occupants' movement among office building rooms. 3 weeks are used for possible frequent occupants' movement and day time iden-

tification, and frequent occupants' movement and daytime intervals selection, and the fourth week is used for the test. As we consider weekdays, and as movement is occupant centric, then we obtain 5 movement graphs for each week and each occupant, which means 16 graphs in total: 12 graphs are used for movement learning, and 4 graphs for the test.

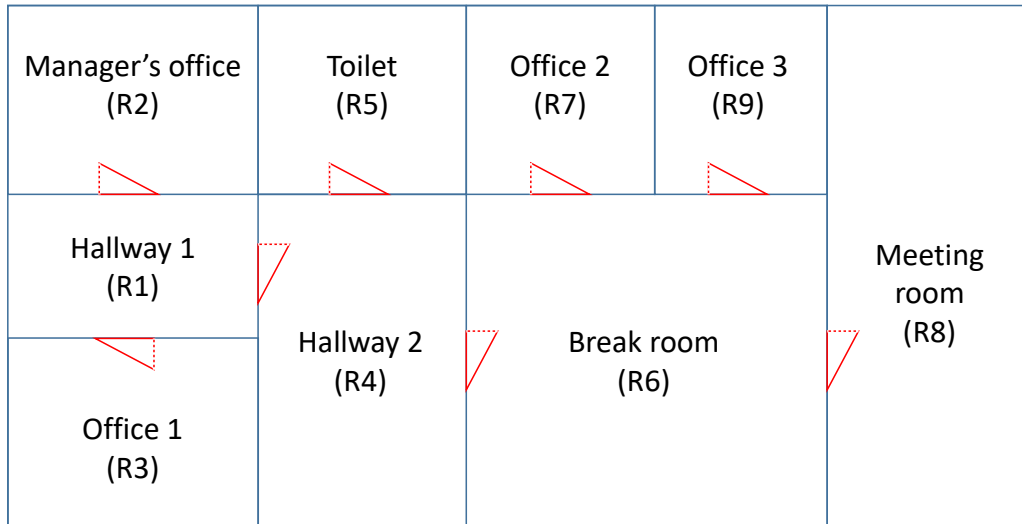
Our synthetic dataset contains the movements of all the occupants among offices of the building. For example, an occupant starts his day around 8 a.m. and works mainly in office room 2 ( $R2$ ). He takes his lunch break around 12h30 p.m. and has a meeting in the meeting room ( $R8$ ) once or twice a week. Occupant gets into/out of the building from hallway 1 ( $R1$ ), since it represents the entrance of the building. Therefore, the day starts when the occupant enters hallway 1 ( $R1$ ) and ends when the occupant leaves the building from this hallway.

#### 4.5.2 Similar occupant movement grouping and daytime identification

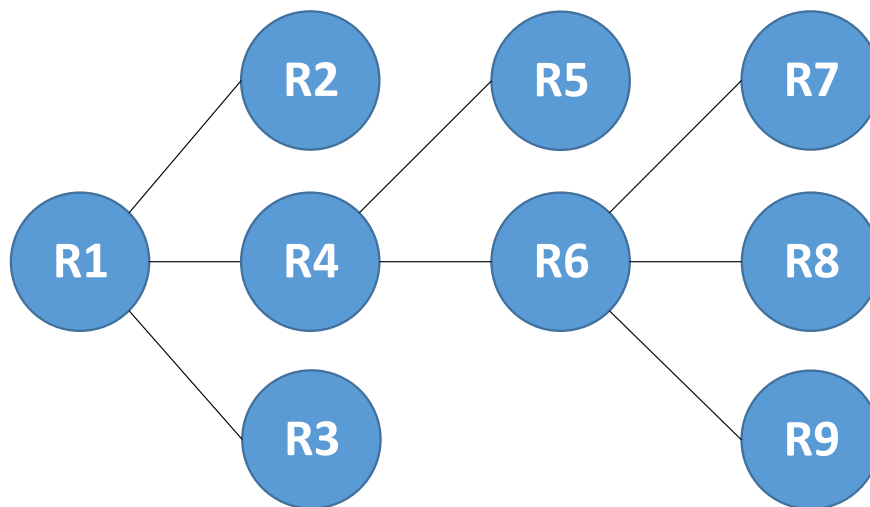
To group similar time movements between 2 rooms  $Rx$  and  $Ry$  we use a hierarchical clustering algorithm. Each movement is represented as follows:  $M(Rx, Ry)_n = [D : HH : MM]$ , where  $x$  and  $y$  represents room source and destination, respectively,  $n$  is the  $n^{th}$  movement from  $Rx$  to  $Ry$ ,  $D$  is the movement day, HH and MM represent hour and minute time information of the movement. For example,  $M(R1, R4) = \{M(R1, R4)_1, M(R1, R4)_2, M(R1, R4)_3, \dots, M(R1, R4)_n\}$  represents the  $n$  possible movements from  $R1$  to  $R4$ , and vice-versa.

This algorithm regroupes these movements. Fig. 4.4 illustrates an example of movement grouping for movement  $M(R1, R4)$ . Each group is in a dotted-line rectangle. For example, movements  $M(R1, R4)_1, M(R1, R4)_3, M(R1, R4)_{10}$  are grouped into the same cluster. Thus, we replace all their timestamps by a single timestamp, which is the minimum timestamp among all these four movements.

We use hierarchical clustering to identify time interval groups of occupant movements



(a) Example of building and its rooms.



(b) Graph-based building and its rooms representation.

Figure 4.3: Example of building.

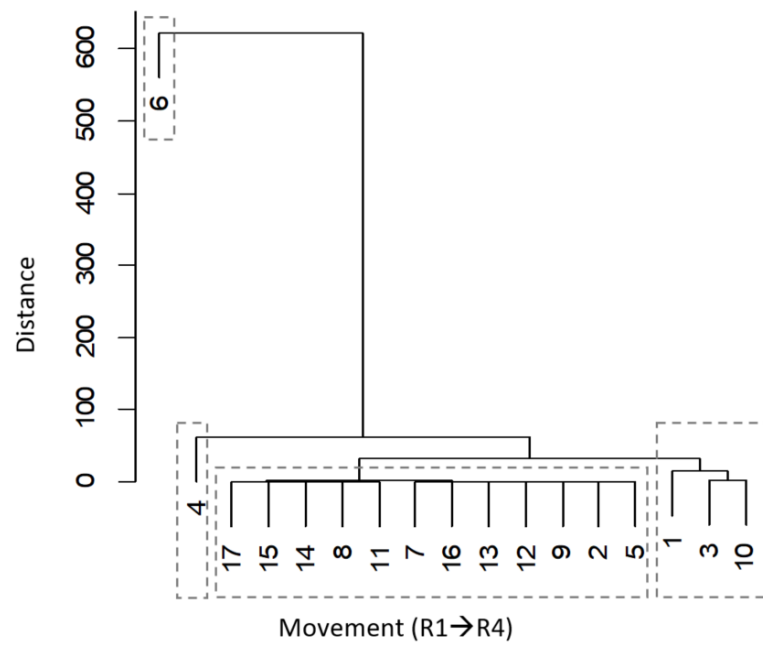


Figure 4.4: Possible movements from R1 to R4.

Table 4.1: Time interval day groups and representative days calculation.

Number of decompositions	Number of days in the decomposition	Distance between day $D_x$ and other days in the same group ( $DS_x$ )	Centroid of group
5	4	$DS_7 = 13.58,$ $DS_{12} = 13.58$	$D_7$
7	34	$DS_2 = 8.66, DS_{14} = 8.66$	$D_2$
9	17	$DS_1 = 8.74$ $DS_4 = 8.30, DS_5 = 7.26$ $DS_6 = 4.54, DS_8 = 13.94$ $DS_{13} = 6.88, DS_{15} = 6.37$	$D_6$
11	5	$DS_2 = 144.63$ $DS_2 = 75.45,$ $DS_2 = 74.26$	$D_3, D_{10}, D_{11}$

for each day of the three first weeks. The results show that there are four possible decompositions: 1) 5 subintervals, 2) 7 subintervals, 3) 9 subintervals 4) 11 subintervals. Fig. 4.5 illustrates an example of day  $D_1$  decomposition. It shows that  $D_1$  movements can be split into 7 subintervals, such that 4 subintervals correspond to occupant's movements and 3 subintervals correspond to no movement periods (occupant stays in the same room). Table 4.1 shows day groups, the number of days that have the same decomposition, distance calculation, and identified representative days. We set  $\varepsilon = 30$  for distance calculation. Thus, days  $D_3$ ,  $D_{10}$ , and  $D_{11}$  are not unified since the distance among them is greater than  $\varepsilon$ . For  $D_7$  and  $D_{12}$  (resp.  $D_2$  and  $D_{14}$ ) the distance between them is less than  $\varepsilon$ . Thus, each of them can represent the other. We have chosen  $D_7$  and  $D_2$ . Since  $D_6$  has the lowest distance among days  $D_1, D_4, D_5, D_6, D_8, D_{13}, D_{15}$ , which is less than  $\varepsilon$ , then  $D_6$  represents the centroid of this group. As it is shown in Table 4.1, only 4 and 5 days are suitable for decomposition to 5 and 11 subintervals, respectively, but these decompositions are not representative. Therefore, we only consider the possibilities of 7 and 9 decompositions. Moreover, decomposition to 9 subintervals contains subintervals of 16 min, 6 min, and another of 9 min, which is too narrow to be used in the heating/cooling decision-making process. Consequently, decomposition to 9 subintervals is also discarded.

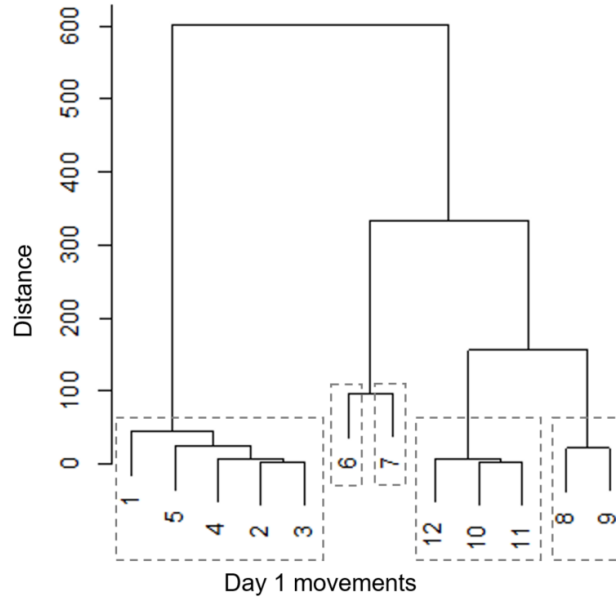


Figure 4.5: Day 1 intervals decomposition.

### 4.5.3 Frequent Movement Sequence Identification

As shown in Subsection 4.5.2, we decompose the daytime interval into 7 subintervals. Subgraphs of the same subinterval are grouped. At the end of the process, we obtain 7 groups of 12 subgraphs (a subgraph for each occupant for each day of weekday). We used a graph mining algorithm (gSpan) to identify frequent occupants' movements among building rooms. The fourth week of the dataset is used for the test. Its days are divided into 7-time intervals as day  $D_7$ , and then compared to frequent occupant movements given by  $D_7$  decomposition.

### 4.5.4 Results

Fig. 4.6 (from Fig. 4.6a to Fig. 4.6d) displays the comfort computed for each occupant during the test week. The comfort is the proportion between the duration where the occupant is in a heated/cooled room and the total occupancy duration. We notice that our system ensures an average occupant's comfort from 47.78% (Fig. 4.6d) to 88.01%

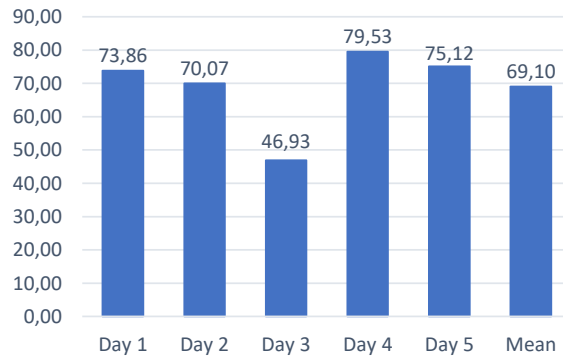
(Fig. 4.6c). We aggregated the comfort of the test week for the occupants, in addition to the saved energy (in percentage compared to the continuous heating/cooling method) in Fig. 4.7. Total comfort provided is up to 77.82%, with an average of 71.28%. The energy saved by our approach is up to 62.21%, with an average of 58.16%.

## 4.6 Summary of the Results

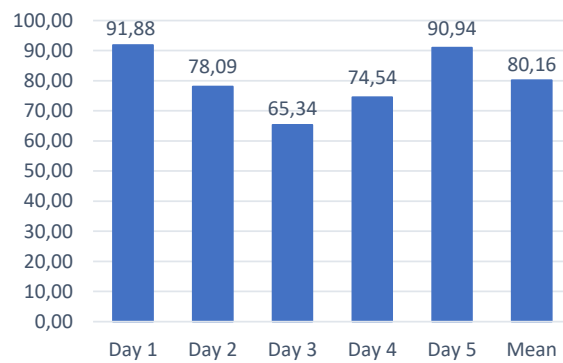
In the previous section, we have evaluated the performance of the proposed approach. Simulation results showed that:

- Our approach helps model and understand occupants' movement within buildings using graph abstraction. This abstraction level can help represent any type of buildings, whatever its dimension.
- By using our solution, we can help minimize the energy consumed by the HVAC system by up to 62.21%. This represents an enormous energy economy for the building because heating and cooling represent the greatest energy consumption of the buildings.
- The average comfort provided by our system is on average 71.28%. We notice that more than 25% of the occupants' comfort is wasted and this is one of the disadvantages of this method. Nevertheless, we have considered as a moment of discomfort any moment the occupant is within a room/hallway which is not heated, even the moments when the occupant walk in an unheated hallway, to go from a room to another, but if we have considered that the shortest occupancy duration (for example less than 10 minutes) of a room does not impact the comfort of the occupant, then the comfort going to be higher.

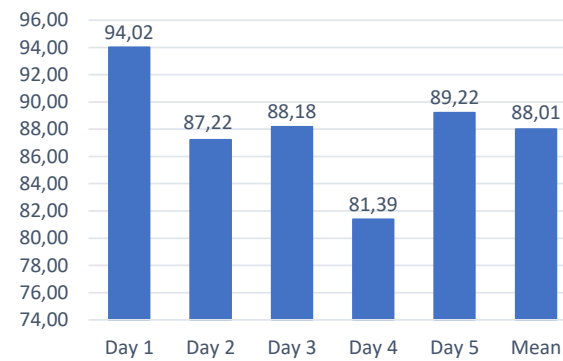




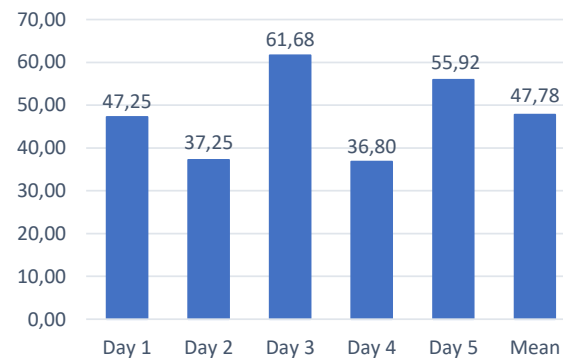
(a) Comfort computed for occupant 1 in test week.



(b) Comfort computed for occupant 2 in test week.



(c) Comfort computed for occupant 3 in test week.



(d) Comfort computed for occupant 4 in test week.

Figure 4.6: Provided comfort for the four occupants in each day of the fourth week, plus the mean comfort for each occupant.

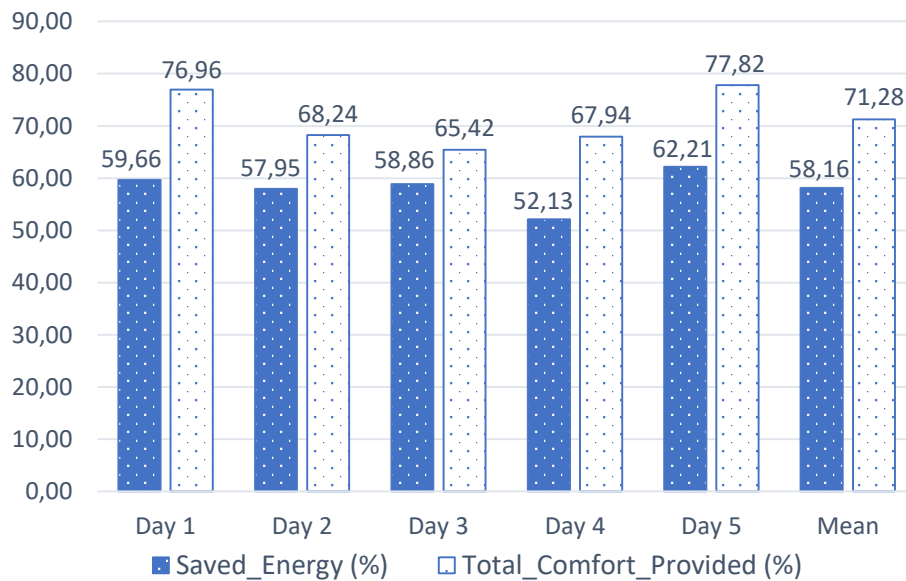


Figure 4.7: Daily saved energy and provided comfort.

## 4.7 Conclusion

Building occupant behavior is an important factor that impacts building occupant comfort and energy consumption, especially by HVAC systems. In this chapter, we studied occupant movement behavior based on a temporal graph model, such that nodes represent building rooms and edges represent occupants movement among building rooms. The main contributions of this chapter can be summarized as follows:

- Temporal graph-based building and occupant movement representation: The use of temporal graphs gives a higher abstraction level of any building types (habitats, offices, etc.), which helps apply many graph theoretic approaches to extract information about occupant's movement habit and, then construct a realistic prediction model of the occupant interaction with the appliances, light, HVAC system, etc. in order to optimize the energy consumption.
- Data Mining algorithm-based rooms-changes and time-frames patterns detection: We made use of hierarchical cluster analysis to learn temporal occupants movement

information, which helps identify the time intervals in which the occupants visit a subset of specific rooms. The use of this technique helps avoid empirical time interval decomposition and therefore avoids possible mistakes.

- Graph Mining-based occupant behavior prediction: We make use of a graph mining algorithm to identify frequent occupants' movements. The graph-based representation helps us detect the rooms' relationship based on the occupant movement habits. Thanks to GM algorithms we can extract the occupant behavior pattern and then build an accurate prediction model that helps predict the rooms that the occupant may visit during a future time interval and then, anticipate some decisions to avoid impacting the occupant comfort, such as rooms temperature, and building energy consumption, such as turning-off HVAC systems in the rooms foretasted to remain unoccupied rooms.

To gather necessary data for this approach we need to track the movement of each occupant of the building, from the moment he/she gets into the building, until the moment he/she leaves the building. Although, this approach helps deeply understand the exact movement behavior of each occupant of the building. This solution has some disadvantages, which are:

- Privacy problem: Tracking occupant's movement within a building represents an invasion of the privacy of the person, which may psychologically affect the occupants' real movement behavior through the building. This is the main weakness of our approach.
- Deploying a tracking system requires a high-quality localization detection system. Therefore, any wrong detection of the occupant localization (occupant is in the room  $R_1$  but the system detects that he/she is in the room  $R_2$ ) can affect the accuracy of the prediction model.

To fulfill the localization problem, there are some solutions in the literature [96, 97]. Nevertheless, the problem of privacy is still partially resolved. Furthermore, sometimes

---

the model can give come wrong prediction that impact the occupants' comfort. Hence, these issues need to be fixed. Therefore, we need to propose a solution that is less intrusive for the occupant's privacy, but at the same time needs to be accurate. In fact, in chapter 3 we have proposed a room-centric approach that does not require information about each building's occupant. As such, it protects occupant privacy. Nevertheless, this approach does not consider the occupant behavior but only consider rooms occupancy pattern. Therefore, it needs to be improved. Besides, in this chapter (chapter 4), we have proposed an approach that takes into account the occupant movement behavior and occupancy preference in a deep way, but, it represents a real privacy problem. Consequently, we need to propose an intermediate solution that is adaptable to occupant behavior and occupant behavior habit changes, and, at the same time, gives an accurate prediction model. This will be the focus of the next chapter.



## Chapter 5

# Occupant Behavior Prediction and Real-Time Correction-based Smart Building Energy Optimization

### Contents

---

<b>5.1</b>	<b>Introduction</b>	<b>86</b>
<b>5.2</b>	<b>Context and Motivations</b>	<b>87</b>
<b>5.3</b>	<b>Related Work</b>	<b>88</b>
<b>5.4</b>	<b>Methodology</b>	<b>91</b>
5.4.1	Occupant Movement Prediction (OMP)	91
	Time interval Occupants Movement timestamp Unification (TOMU)	92
	Frequent Movement sequence time interval Identification (FMI)	92
	Predicting building occupants' movements	93
	Decision making	94
5.4.2	Real-time Occupancy prediction verification and Correction (ROC)	95
	Real-time decision correction	96
	Occupancy prediction correction	96

<b>5.5</b>	<b>Evaluation and results</b>	<b>97</b>
5.5.1	Dataset description	97
5.5.2	Experiment description	99
5.5.3	Result and discussion	100
<b>5.6</b>	<b>Summary of the Results</b>	<b>102</b>
<b>5.7</b>	<b>Conclusion</b>	<b>102</b>

---

## 5.1 Introduction

In the previous chapter, we focused on graph-based building energy optimization. We have proposed a model that gives a higher abstraction representation of building and its occupants, and learns occupant movement behavior within the building. This model gives a close vision of occupant’s movements habit hence it can give a clear image of energy that can be saved. Nevertheless, this model is occupant-centric, and requires tracking each occupant’s movements, from the moment he/she enters the building until the moment he/she leaves it. This can be seen as a privacy intrusion by many occupants. Furthermore, this model can give wrong predictions and, without correcting it, the model may cause lack of occupant’s comfort, in false-negative cases. For example, if during the  $t^{th}$  time-interval, a room is predicted as unoccupied, the HVAC will be turned OFF during the predicted time-interval. But, if during this time-interval, an occupant gets into the room, his/her comfort will be reduced. Consequently, we need to cope this problem. Nevertheless, we have presented, in chapter 3, a real-time occupancy detection system, that can provide room occupancy state using real-time collected data.

To alleviate the two above mentioned inconvenient we propose, in this chapter, to combine the model presented in chapter 4 with the model presented in chapter 3. We aim at predicting occupants’ movements among rooms and use the predicted movements to deduce room and space occupancy using the prediction model from chapter 4 in the building. The latter is then used to preheat/pre-cool rooms. However, since prediction

models are not always that accurate, it is possible to face situations where HVAC of some rooms are activated while these are empty or vice-versa, leading to either a waste of energy or a lack of occupant's comfort. To deal with this issue, we make use of sensors to detect real-time rooms occupancy state (inspired from chapter 3) to correct the prediction when necessary. To achieve this, we developed a graph mining-based optimization approach that combines occupant behavior prediction and a real-time correction.

The remainder of this chapter is organized as follows. Section 5.2 presents the context and motivation. Section 5.3 discusses related work. Section 5.4 introduces our methodology. Evaluation settings and results are discussed in section 5.5. Discussions are given in Section 5.6. Section 5.7 concludes the chapter and proposes future directions for future work.

## 5.2 Context and Motivations

In any occupied building, rooms are interconnected through doors or hallways. Occupant's moves within rooms and create a spatiotemporal relationship between building rooms. To understand and learn occupant's movements behavior within the building, we can use these movements traces to build a pattern of movement behavior that can help predict occupant's behavior and then, optimize energy (example: turn-off HVAC system in the rooms predicted to be unoccupied in the future) and increase comfort. In this manner, we will mainly focus on the occupant, creating a method that is occupant-centric. Nonetheless, as showed in chapter 4, we need to trace each occupant's movements within the building during all the time the occupant is inside the building.

To address this inconvenience, we develop, in this chapter, a combined approach to optimize building energy usage while ensuring occupant comfort. First, we exploit our graph mining-based prediction model, proposed in [3], to predict room occupancy based on building occupants' movement over time. This occupancy prediction is then used to schedule the functioning of the HVAC inside rooms according to a prediction of their



future occupancy status. Second, we track the occupancy status based on real-time data collected by sensors deployed inside the building. The objective of this second step is to be able to use real-time sensor-based occupancy detection, as proposed in [2], to correct room occupancy prediction errors by reactivating (resp. deactivating) HVAC systems in occupied (resp. unoccupied) rooms, where their occupancy status have been miss-predicted by the prediction system. The main objective is still to decrease energy consumption while ensuring building occupants' thermal comfort. We experimented our approach based on a realistic building occupancy dataset and the result shows that our combined approach allows saving up to 39.09% of HVAC energy consumption while ensuring up to 99.39% of thermal comfort for building occupants. Furthermore, while we have collected occupants' movement data only for a specific duration but not after it, we cannot know the occupancy behavior of each occupant, but the system only gives a prediction of room occupancy state. Hence, there will be no occupant's tracking and thus, occupancy is preserved.

### 5.3 Related Work

Energy is of vital importance in all sectors such as transport, industry, households, and services. Under the new policy scenario, global energy consumption and *CO2* emissions are expected to increase by approximately 50% in 2018 to 19.8% 2050, respectively [98]. In this context, buildings consume a huge amount of energy reaching about 40% of the EU and US residential and commercial buildings [99]. Therefore, optimizing buildings' energy consumption becomes a big challenge and an important topic to study not only for researchers but also by industries and public authorities as well.

Building energy consumption depends on both external factors, such as buildings' envelope and orientation [100, 101], and internal factors seen as indoor environment conditions such as occupancy and occupants' behaviors [102, 103, 104]. Recent research results showed significant differences between simulated energy consumption and actual energy consumption. Furthermore, they showed why most traditional building systems,

that usually apply a fixed operating schedule according to certain regulations (such as ASHRAE or user surveys), results in wasted energy and discomfort for the occupants [85].

Despite the above mentioned findings as well as a lot of research in the field of building systems, there is a lack of research work on occupant behavior detection and analysis [102]. Building occupants highly impacts HVAC energy consumption, which represents over 50% of typical building's energy consumption [105]. Therefore understanding building's occupancy and occupant's behavior can help optimize HVAC energy consumption [69, 106, 107].

Many researches in the field of sensor-based energy optimization in buildings have focused on building occupants' behavior, room or space occupancy status, and occupancy prediction. Several data-driven methods have been introduced encompassing optimization approaches, probabilistic approaches, and machine learning approaches, to name a few.

Kusiak et al. [108] have considered data-driven optimization methods to reduce HVAC energy consumption in offices by adjusting control settings (supply air static pressure and supply air temperature). They tested eight supervised machine learning algorithms to simulate the nonlinear relationship between controlled settings, energy consumption, and uncontrolled variables. They found that, among the tested algorithms, multiple-linear perceptron (MLP) provides the highest accuracy. Their system saves up to 7% of the HVAC energy consumption. Brooks et al. [109] have proposed an improved Variable Air Volume (VAV) HVAC system based on occupancy. Their results demonstrates from 29% to 80% of energy reduction in 5-room building served by the improved VAV HVAC system. Nevertheless, the HVAC equipment was common to all these rooms, which prevents to finally control the HVAC system for each room independently.

Chen et al. [110] have proposed a stochastic occupancy modeling based on a Markov chain. Two new non-uniform Markov chain models have been proposed, which have been harnessed in two situations: multi-person single-zone (MOSZ) and multi-person multi-

zone (MOMZ). The novelty of this study is that the state of the Markov chain is has defined as the increment in the number of occupants, not the number of MOSZ occupants. For MOMZ, the state of the Markov chain is a vector where each component is the occupancy increment in each region. This is performed with the aim to reduce, the computing burden. To compute the probability matrix, a maximum likelihood estimation is used. The approach was compared to the ABM model proposed in [111] and the results showed that it performs better, under some assumptions. A real-time occupancy prediction method based on spatio-temporal history analysis is proposed by Adamopoulou et al. [89]. In this study, the data is collected in real time using depth image cameras as well as acoustic and motion sensors. Their goal is to calculate the occupancy rate prediction based on the collected data. Yilmaz et al. [82] have developed a bottom-up random occupant behavior modeling method to predict the use of household appliances in home buildings. The study is limited to home buildings to determine data that affect energy consumption. The above mentioned studies did not consider building rooms occupants' movements behavior relationship, which is an important factor that may cause changes in the occupancy status of the building rooms. In fact, rooms occupancy state is related to occupants' movements among rooms. Indeed, knowing the occupant movement pattern, we can predict the rooms that he would visit, and the ones that he would not visit, and then anticipate the activation (resp. deactivation) of the HVAC in the rooms accordingly, in order to improve his comfort and minimize energy consumption.

Related works addressed either rooms occupancy or data-driven energy efficiency, but not both at the same time. Additionally, from prediction error detection and correction viewpoint, to the best of our knowledge, there is no work in the literature where that issue has been tackled in the field of building energy consumption. Therefore, in order to achieve the objectives of our smart building energy optimization, we describe in the subsequent sections our graph mining-based approach combined to an optimized method to deploy sensors inside a building to collect data for a real-time room occupancy status determination and erroneous occupancy prediction detection and correction.

## 5.4 Methodology

Our approach consists in two main steps: (i) Occupant Movement Prediction (OMP) to predict occupant movements, and HVAC system control plans (detailed in Subsection 5.4.1), and (ii) Real-time Room Occupancy status Correction (ROC) to correct prediction (detailed in Subsection 5.4.2).

### 5.4.1 Occupant Movement Prediction (OMP)

Occupant Movement Prediction (OMP) aims to compute future occurrences of occupants' movements within building rooms, and predict the occupancy status of the rooms, based on occupants' historical movements. Anticipating movements and occupancy allow a pre-commanding the HVAC system to pre-heat/pre-cool rooms before occupant arrival.

Let  $B$  be a building of  $N$  rooms, which are connected with hallways and corridors. Occupants can move from a room to another. We represent their movements by a graph  $G(V, E)$ , such that  $V$  represents the building rooms, and  $E$  represents possible movements among these rooms.  $O$  represents an occupant of  $B$ .  $M_V$  is the matrix of all occupant's movements within  $TI_i = [t_1, t_2]$  time interval, such that:

$$M_V = V_{[t_1, t_2]} \times V_{[t_1, t_2]} \equiv (m_{i,j}) = \begin{cases} 1 & \text{if } (V_i, V_j) \in E_{[t_1, t_2]} \\ 0 & \text{otherwise.} \end{cases} \quad (5.1)$$

As we suggested in [3], a graph-mining approach is a suitable approach for predicting  $M_V$ . The whole process consists in the following steps:

- Time interval Occupants Movement timestamp Unification (TOMU)
- Frequent Movement sequence time interval Identification (FMI)
- Building occupants' movements prediction
- Decision making

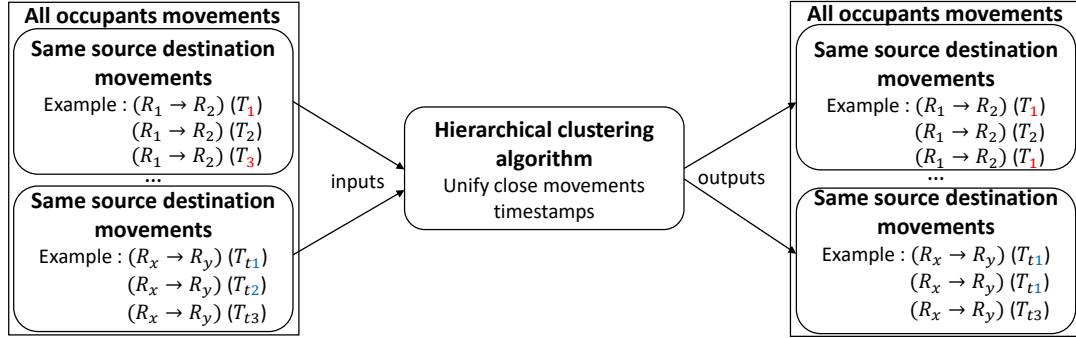


Figure 5.1: TOMU algorithm unifies timestamps  $T_1$  and  $T_3$  to become  $T_1$  for two different movements, as well as for  $T_{t1}$  and  $T_{t2}$ , which are unified as  $T_{t1}$ .

### Time interval Occupants Movement timestamp Unification (TOMU)

Graph mining algorithm is an accurate enough approach in selecting frequent subsequences (in our case: subgraphs of movements). It distinguishes even the same movements happens within very close times. As we are dealing with human movements, the distinction between such movements is not that relevant, we propose to unify them with the aim to accelerate the algorithm and to provide human readable results. To perform this timestamps unification, we propose to use a hierarchical clustering algorithm that groups movements with close timestamps then, we give a fixed timestamps (the smallest timestamps in the group) to each movement in the same group. Figure 5.1 illustrates the TOMU process through an example of timestamp unification of close movement timestamps.

To avoid manually unifying timestamps, we implemented the algorithm by using a hierarchical clustering.

### Frequent Movement sequence time interval Identification (FMI)

To optimize prediction periods, we need to divide a day (24 hours) into a Time Interval List  $TIL = \{TI_1, \dots, TI_t\}$  of  $t$  time intervals, where  $TI_{i,i \in \{1..t\}} = [t_{S_i}, t_{E_i}]$ , with  $t_{S_i}$  is the beginning of the interval, and  $t_{E_i}$  is the end of the interval.

We apply on each day  $d$  a hierarchical clustering method to generate its set of time intervals, denoted by  $TIL_d$ , then we eliminate all  $TIL_d$  that contain at least a  $TI_i$  such that  $TI_i < minTime$ , with  $minTime$  is the required duration to heat/cool a room to the comfort temperature.

A second elimination process is required to select, from obtained  $TIL$  lists, the representative one, i.e. the one which is the most frequent. It is then considered the  $TIL$  decomposition for all days.

### Predicting building occupants' movements

To proceed with the occupants' movements prediction. We need to rely results on the two previous steps (TOMU and FMI). We divide occupants' movements of each day into groups based on the timestamp of each movement. We denote by  $M_v = \{M_{vD1}, \dots, M_{vDd}\}$  the list of movements of all days, where  $M_{vDi}$  consists of all the movements of day  $Di$ , such that each movement is represented by the triplet source room ( $R_s$ ), destination room ( $R_D$ ), and timestamp ( $t_{R_s, R_D}$ ) representing the time when the movement happened. Therefore,  $M_{vDi}$  of  $m$  movements is defined as:

$$\begin{aligned} M_{vDi} = \{E_1 = [R_{s1}, R_{D1}, t(E1)], \dots, \\ E_m = [R_{sm}, R_{Dm}, t(Em)]\} \end{aligned} \quad (5.2)$$

Algorithm 2 describes the sub-movement list extraction for each  $TI$  of  $TIL$ . The result of this algorithm is then used in a second step by the graph mining algorithm to extract, for each  $TI$  of  $TIL$ , occupants' movements prediction.

---

**Algorithm 2** Sub-movements selection algorithm.

---

**Require:**  $MT = \{M_v(D_1), \dots, M_v(D_d)\}, TIL = \{TI_1, \dots, TI_t\}$

**Ensure:**  $M_v(TI_1), \dots, M_v(TI_t)$

**BEGIN**

**for each**  $TI_i$  **in**  $TIL$  **do**

**for each**  $M_{vD_i}$  **in**  $M_v$  **do**

**for each**  $E$  **in**  $M_{vD_i}$  **do**

**if**  $t(E) \in TI_i$  **then**

                add  $[E, t(E)]$  to  $M_v(D_i)_{TI_i}$

**end if**

**end for**

        add  $M_v(D_i)_{TI_i}$  to  $M_v(TI_i)$

**end for**

**end for**

**return**  $M_v(TI_1), \dots, M_v(TI_t)$

**END.**

---

### Decision making

After extracting list of movements of each day and for each  $TI$  of  $TIL$ , we compute the movement prediction matrix as follows:

$$M_P = V_{[t_1, t_2]} \times V_{[t_1, t_2]} \equiv (m_{i,j}) = \begin{cases} 1 & \text{if } (V_i, V_j) \in E_{[t_1, t_2]} \\ 0 & \text{otherwise.} \end{cases} \quad (5.3)$$

We extract, from  $M_P$ , the Occupancy Prediction vector, denoted by  $PO_{[t_1, t_2]}$  for Predicted Occupancy during  $[t_1, t_2]$ , of  $V$  elements, such that each room defined as a destination is considered as an occupied room. More precisely,  $PO_{[t_1, t_2]} = [P_{1s}, \dots, P_{vs}]$ , such that  $P_{xs}$  is the predicted occupancy status of room  $x$ , computed using Equation 5.4, with  $s$  is the source room, and  $v$  is the number of rooms in the building.

$$P_{sx} = \begin{cases} 1 & \text{if } \exists (V_i, V_x) \in E_{[t_1, t_2]}, i \in [1, v] \\ 0 & \text{otherwise.} \end{cases} \quad (5.4)$$

We introduce our Occupants Movements' Prediction algorithm (OMP), listed in Al-

gorithm 3, in order to compute the occupancy prediction for all rooms during all the considered prediction periods.

---

**Algorithm 3** OMP algorithm.

---

**Require:**  $MT = \{M_v(D_1), \dots, M_v(D_d)\}$

**Ensure:**  $TIL, M_v, PO_{TIL}$

**BEGIN**

$TIL \leftarrow$  Generate time interval list using hierarchical clustering

**for each**  $TI$  of  $TIL$  **do**

$M_v(TI) \leftarrow GM(MT, TI)$

$PO_{TI} \leftarrow$  Generate Prediction vectors from  $M_v$  using Graph Mining Algorithm

**end for**

**return**  $TIL, M_v, PO_{TIL}$

**END.**

---

Based on Algorithm 3, the rooms are predicted as occupied during  $TI_i = [t_a, t_b]$  will be pre-heated/pre-cooled  $t$  minutes before  $t_a$ . As this process is dependant on the human movements which may change transiently, we added in our next step the real-time occupancy verification, and prediction correction.

#### 5.4.2 Real-time Occupancy prediction verification and Correction (ROC)

Relying on the occupants movement prediction, the system decides about the HVAC status for each room of the building during each time interval  $TI = [t_S, t_E]$  of  $TIL$ . As such prediction may contain errors, we adjust to it a new algorithm called ROC for Real-time Occupancy prediction verification and Correction. The real-time room occupancy status at a timestamp  $t$ , denoted by  $RO_t$ , is computed based on data collected from the sensors deployed inside the building as detailed in [2].

$RO_t = [R_1, \dots, R_v]$ , such that  $R_x$  is the occupancy status of room  $x$  at time  $t$ .

We compare  $PO_{TI}$  with  $RO_t$  to detect occupancy prediction errors. The comparison result is denoted by  $OC$  for Occupancy Comparison and is computed by equation 5.5. As  $PO_{TI}$  and  $RO_t$  are 2 vectors of binary values, then  $OC$  is the vector containing the absolute values of the subtractions between these 2 vectors. It is also a vector of binary



values.

$$OC = [OC_1, \dots, OC_v] = |PO_{TI} - RO_t| = \quad (5.5)$$

$$[|PO_{TI}(R_1) - RO_t(R_1)|, \dots, |PO_{TI}(R_n) - RO_t(R_n)|]$$

The obtained vector  $OC$  is used for correcting predicted decisions, and occupancy prediction (for new time interval  $[t, t_E]$ ). The two corrections are Real-time decision correction and Occupancy prediction correction, detailed below. The ROC algorithm is summarized in Algorithm 4.

### Real-time decision correction

Each  $OC_i$  of  $OC$  is interpreted as follows:

- If  $OC_i = 1$  then it means that the value of  $PO_{TI}$  and  $RO_t$  are different, and the real occupancy status is different from prediction occupancy status for room  $i$ . Therefore, the control system will reverse the HVAC status of that room: if the HVAC is turned ON (resp. OFF), the system will turn it OFF (resp. ON).
- Otherwise, occupancy prediction of room  $i$  is considered as correct. No change is required for the HVAC status in room  $i$ .

### Occupancy prediction correction

After the real-time occupancy detection at time  $t$ , the time interval  $[t_S, t]$  is no more valid. Therefore, we need to generate new occupants' movements and occupancy prediction for new time interval, denoted by  $NTI = [t, t_E]$ . Thus, new  $M_v(NTI)$  must be generated. Furthermore, some  $M_v(D_i)_{NTI}$  should be eliminated. We eliminate all  $M_v(D_i)_{NTI}$  such that  $PO_{NTI}(M_v(D_i)_{NTI})$  is different from  $RO_t$ . The new  $M_v(NTI)$ , denoted by  $NM_v(NTI)$ , is used to generate both new prediction movement matrix and vector using graph mining algorithm the same way we have done for the precedent inter-

val.

**Example.** Suppose that we predict occupancy for time interval [10:00, 12:00] and we use occupancy detection each 30 minutes  $t \in \{10:30, 11:00, 11:30\}$ . If, for  $t=10:30$  the system detects that  $OC_i = 1$ , then the system will generate new occupant's movement and occupancy prediction for the new time interval started from  $t$ : [10:30, 12:00]. The same tests and decision will be made for  $t=11:00$  and  $t=11:30$ .

---

**Algorithm 4** ROC algorithm.

---

**Require:**  $M_v TI, TI, PO_{TI}, RO_t, MT = \{M_v(D_1), \dots, M_v(D_d)\}$

**Ensure:**  $RO_t, NM_V TI$

**BEGIN**

$OC \leftarrow |PO_{TI} - RO_t|$

$NM_v \leftarrow (MT/OC)$  {selecting from MT only elements that verify  $RO_t$ }

$NM_v(TI/OC) \leftarrow GM(NM_v, TI)$  {generate new  $M_v$ }

**return**  $RO_t, NM_V TI$

**END.**

---

## 5.5 Evaluation and results

In this section, we detail the evaluation carried out to assess our solution, and we discuss the obtained results.

### 5.5.1 Dataset description

We have considered a one-month set of observed movements of four persons within university building rooms. The first three weeks are used for model training: Time interval Occupants Movement timestamp Unification (TOMU), Frequent Movement sequence time interval Identification (FMI), and occupants' movements prediction model construction. The fourth week is used for Occupant Movement Prediction (OMP), OMP performance verification, and Real Time Occupants' Movements Correction (ROC).

Figure Fig. 5.2 illustrates the rooms of the considered building. Doors are represented

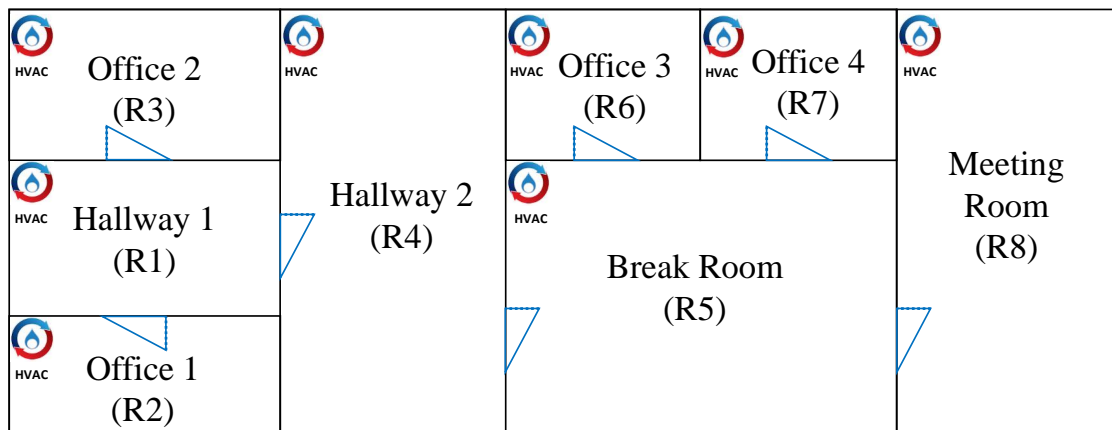


Figure 5.2: Building Rooms representation.

by blue triangles, Each room and space is control by an HVAC system, as represented on the top left corner of each of these.

We represent the building as a graph, such that building rooms are the set of vertices or nodes, and possible movements within these rooms correspond to the edges. Figure 5.3 illustrates the graph drawn from our building (Fig. 5.2).

The data used in the simulation gathered the observed movements within the building of the four occupants considered. For example, occupant  $O_1$  mainly works in Office 1. He/she enters the building at around 8 : 30 *am*, works in office 1, takes his/her lunch at the Break Room at around 12 *pm*, has a meeting once or twice a week in the Meeting Room, and leaves the building at 6:00 *pm*.

We start recording movements from the moment when occupants enter the building; in our case, Hallway 1 ( $R1$ ) represents the entrance of the building. The last recorded movement is by the end of the day or when the occupants have left the building. Therefore, each movements graph depends on a different occupant's behaviour. Hence, in our case, we can end up with four movements' graphs per day: a graph for each occupant.

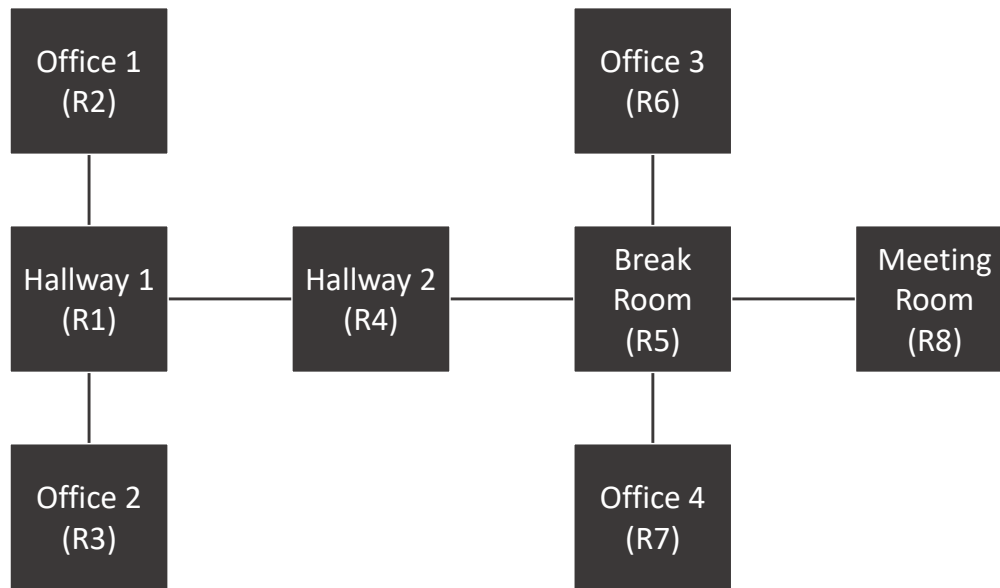


Figure 5.3: Graph-based Building representation.

### 5.5.2 Experiment description

In Occupant Movement Prediction (OMP) step, we used gSpan (graph-based substructure pattern mining) algorithm, for selecting frequent subgraphs from movement graphs. It requires an important parameter called minimum frequency of acceptance of subgraph (the number of times a frequent pattern has to appear in the graph), which means that the algorithm selects only subgraphs that occur with a minimum number of times greater than or equal to that frequency of acceptance.

In our experiment, we have considered a frequency of acceptance of 20%. Thus, all subgraphs that appear less than 20% for the OMP process are eliminated. This frequency of appearance can seem to be low but it is adequate to our application context since each occupant has its own office room and we need to capture in our prediction model his/her individual movements too. As we have experimented, a higher frequency of acceptance (60% for instance) generates weak model in terms of accuracy of rooms occupancy status, since it limits the prediction model to only the movements occurring among shared spaces in the building such as hallways, corridors and break rooms. Besides, it is worth noticing

that this value is empirical at this level of the experiment, and we need to tune it to obtain a better prediction accuracy.

In Real-time Occupants' movements Correction (ROC) process, we need to define a time period in which we verify room occupancy status. In our experiment, we have compared three time periods: 10 minutes, 20 minutes and 30 minutes.

We computed saved energy percentage by the HVAC system compared to a conventional use of such systems where the HVAC is ON during office hours (from 08h30 *am* to 06 *pm*).

We also computed occupant's comfort percentage. We start with computing comfort duration. It is the cumulative addition of minutes where an occupant is inside a given room where the HVAC system is turned ON. Occupant's comfort percentage is the ratio of comfort duration to total day occupancy duration of the building, for each occupant of that building.

### 5.5.3 Result and discussion

Figure 5.4 displays the percentages of saved energy within all building rooms, during the test week while using OMP. Saved energy is the difference between HVAC activation duration in regular system (turned ON during all working time), and the controlled by our proposed system, i.e. the HVAC is turned ON only on predicted occupancy periods.

Results show that, in our case and with 20% of prediction accuracy, we can save up to 39.09% of HVAC energy consumption, with an average of 38.83%. This result can vary depending on building type and occupants' behavior. Figure 5.5 shows provided occupants' comfort comparison between OMP and ROC using its 3 verification time periods (10 minutes, 20 minutes and 30 minutes). It shows that, in our case, OMP provides between 59.85% and 73.16% of occupants' comfort when using OMP only. When ROC is also deployed, comfort can reach 99.39%, 98.86% and 98.93% for 10 minute-, 20 minute- and 30 minute- time period, respectively. This high difference between OMP and

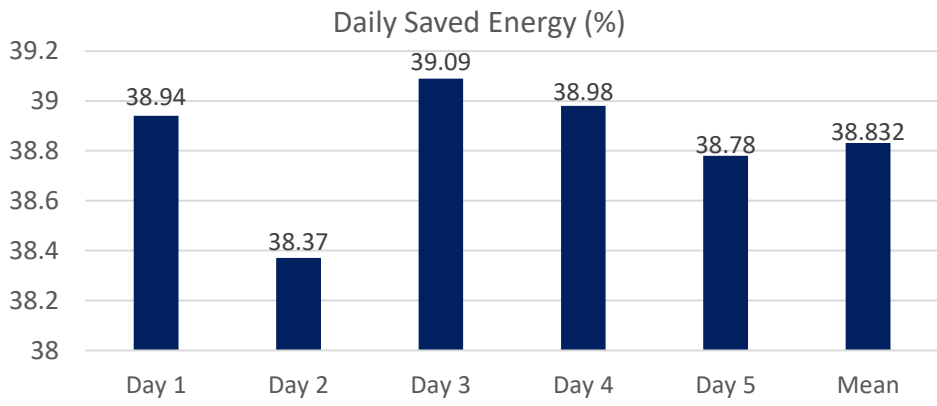


Figure 5.4: Percentage of daily saved Energy with 20% of prediction accuracy.

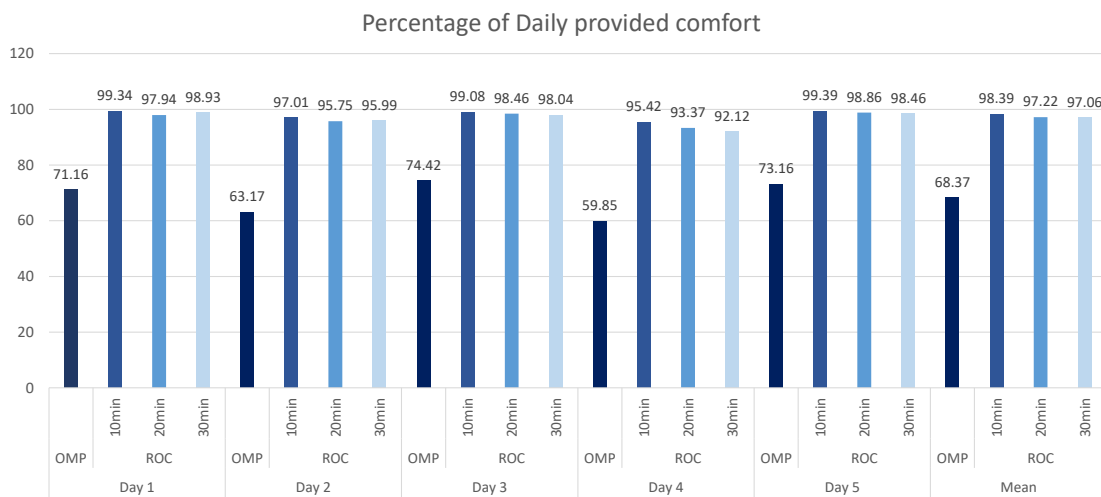


Figure 5.5: Occupants' comfort satisfaction level during the test week.

ROC results is due to the real-time comfort correction. It allows to turn ON instantly the HVAC system if a room was predicted as not occupied during a time interval (HVAC turned OFF), but it is detected as occupied by ROC. Therefore, discomfort is reduced to few minutes (in our case 10 minutes, 20 minutes or 30 minutes).

## 5.6 Summary of the Results

The above simulation results show the importance of combining prediction and correction methods. It shows that:

- Spatiotemporal graph-based Occupant's movement prediction is efficient for optimizing building energy consumption. Up to 39.09% of HVAC energy consumption is saved using our system.
- Any changes in occupant's movements behavior can lead to a lack in the prediction model, and influence occupant's comfort. Prediction errors can decrease occupants' comfort by 59.85% to 73.16%. This is why we need to reinforce it.
- Real-time occupancy detection can help correct prediction errors and increase occupant's comfort within the building. In our experiment, we reach up to 99.39% of occupants' comfort within the building.

## 5.7 Conclusion

In this chapter, we developed a combined approach for Occupants' Movements Prediction (OMP) and Real-time Occupants' Movements Correction (ROC). OMP first predicts occupants' movements among building rooms and pre-heat/pre-cool rooms according to their predicted occupancy. Then, to deal with prediction errors, ROC is activated to correct identified erroneous predictions and to control the HVAC system accordingly. Results showed that OMP can help reduce up to 39.09% of the HVAC consumed energy,

---

and provide between 59.85% and 74.42% of occupants' comfort. The combination with ROC allows to reach up to 99.39% of occupants' comfort.

So far, the ROC approach helps reduce occupants' discomfort, but it does not improve OMP prediction model. The approach proposed above can be improved in many ways. For example, develop a tightly coupled prediction-correction model, i.e. including real-time prediction model adaptation (i.e. the correction is implemented at the level of the model itself) could be interesting to investigate, even though the results we obtained that are already highly competitive. Such a new model could be built using either Markov Chains [112] or the AdaBoost algorithm [113].





# Chapter 6

## Conclusion and Perspectives

### Contents

---

<b>6.1 Contributions . . . . .</b>	<b>106</b>
<b>6.2 Perspectives . . . . .</b>	<b>109</b>

---

This dissertation has made a number of contributions towards joint energy and comfort optimization in IoT enabled Smart Buildings. The key contributions of this dissertation are: (i) a room-centric method using Machine Learning algorithms to determine an adequate set of data and data collection frequency for predicting occupant behavior and thus optimize building energy, accordingly; (ii) the use of graph-based building and occupant movement representation to understand, learn and predict occupant's behavior to reach the targeted energy and comfort optimization; and, (iii) the use of a room-centric real-time occupancy detection to correct occupant behavior prediction errors and minimize occupant's discomfort. A discussion about these contributions is given in Section 6.1. Besides, we discuss the possible future work in Section 6.2.

## 6.1 Contributions

Throughout this dissertation, we tried to answer the following inter-related questions: *what data to use?*, *when to collect these data?* *how to model occupant behavior?* *how to use occupant behavior prediction to minimize building energy consumption and maximize occupant's comfort?* and *how to correct occupant behavior prediction errors impacting occupant's comfort?*. In this section, we are going to show how these questions had been answered.

In our first contribution, we have answered the two first questions, namely: *what data to use?* and *when to collect these data?*. To minimize building energy consumption, we need to understand the occupant's behavior, as this behavior represents one of the most important energy consumption factors within the building. To do so, we need to collect data to construct knowledge about this behavior and then, try to make decisions that target minimizing the energy consumed within the building. For example, turning-off the HVAC system in unoccupied rooms. One of the best and efficient ways of collecting data is to deploy sensors within and around the building. Indeed, sensors allow collecting accurate data 24 hours a day, 7 days a week, automatically, without the need for human interaction. Nevertheless, with the explosion of IoT equipment types and numbers, too many sensors types exist and, deploying random sensors or so many sensors costs time and money to buy, install, and to maintain the deployed sensor system.

Therefore, we need to know what sensors to use to have the data that meet our needs. Therefore, we proposed a method that helps carefully define the set of data to use to meet a specific need. Nevertheless, our method will either not be precise or be too costly to maintain if we do not define the adequate frequency of data collection. In fact, many sensors require specifying the period of data collection, such as  $CO_2$ , humidity, temperature sensors, etc. For this kind of sensors, we can configure a sensor to collect data every few seconds, minutes, hours, or days. As a consequence, the "*When to collect data?*" question becomes vital as the shorter data collection frequency is the more energy the sensors consume, and the more maintenance they require, and batteries changing (in

case of using wireless sensors). At the same time, precision increases. So, what is the adequate trade-off?

In this contribution, we aimed to minimize the number of used sensors and maximize data collection frequency while reaching an acceptable precision. We tested and compared many Machine Learning algorithms to define the efficient algorithm that meets our requirements. Results show that it is possible to build an occupancy predictive model with the Random Forest Classifier algorithm having an accuracy of at least 90%, by using 8 sensors collecting data at a 20-min interval, or 5 sensors collecting data at a 15-min interval.

Our second contribution was about answering the question of *how to model occupant behavior?*. In fact, modeling occupant behavior is important to generate an occupant habit prediction system that allows anticipating occupant movements and then, deactivate some building equipment when their use is not necessary, or reactivating them when it becomes necessary. In this contribution, we propose the use of a higher abstraction level, based on the graph, to represent the whole system, building, and occupant, in which building rooms are represented by graph nodes, and possible direct movements between the rooms are represented by graph edges. This representation gave us a macroscopic vision of the occupant behavior within the building and the inter-relations between rooms, based on this occupant behavior.

Based on this abstraction presentation, we generate, Graph Mining algorithms, the occupant's movement model that we use to predict future occupants' movements and then, decide to activate (resp. deactivate) equipment in rooms that the occupant is forecasted to visit (resp. will not visit) at the predicted time of the visit, i.e. in the future. The purpose of this decision process is to minimize building energy consumption while preserving occupant's comfort. A 4-week dataset of 4 occupant movements among office rooms was used to experiment with our approach. The results showed that our model helps minimize energy consumption by up to 62.21% compared to conventional functioning of HVAC systems, and fulfills up to 94.02% of occupants' thermal comfort.

In addition to this usage, one can also note that such generic graph-based representation can be used beyond what use it for in this dissertation. For instance, virtual rooms can be created on the graph to study a future possible physical decomposition of one or many rooms. For instance, if a hallway is composed of more than one HVAC system, and the attendance on one side of it is higher than the other side then, we can decide to turn-off the HVAC system in the less frequented side of the hallway and, to minimize the loss of heat/cool in the heated/cooled side, we can decide to create, for example, a door to separate the two sides. Therefore, the use of this abstraction does not only giving a higher abstraction level, but it can also help propose future building organization ideas.

Nonetheless, even though this contribution allows preserving a good average level of occupant's comfort, it does not allow to fully satisfy them at all times due to prediction errors. Consequently, we propose, in our third contribution, to combine occupant behavior prediction and a real-time correction, using real-time data, collected by the sensors deployed in building's rooms, to correct prediction errors (i.e. false-negative situations). By doing so, and integrating this in our previous graph-based solution, we can achieve a better equilibrium between building energy consumption and building's occupant comfort. We experimented our approach on simulated data and results showed that our model optimizes up to 39.09% of HVAC energy consumption, and provides up to 99.39% of occupants' comfort.

Another feature of our proposed solution is that it allows us to preserve occupant's privacy. In fact, using graph-based occupant behavior prediction requires real-time occupant's behavior tracking, but while using room-based real-time occupant behavior detection method, the system does not require knowing the behavior of each occupant, but only the room occupancy states (occupied or not, which equipment are on/off) is enough to correct prediction errors.

Overall, these contributions give a new vision to occupant behavior understanding in smart building and show the importance of selecting accurate data to have the most precise information. This allows for increasing chances of finding the way to building

energy optimization without sacrificing its occupant's comfort.

## 6.2 Perspectives

The different studies and results described here opened up many lines for future work to explore. We present some of these ideas in the following.

**Real-time prediction correction:** In our last contribution, we proposed to correct the prediction error in a reactive way. In fact, this correction increases the occupant's comfort through a proactive method. Indeed, additional improvements can be achieved by changing the prediction model based on detected prediction errors, and thus providing an adaptive dynamic prediction system that takes naively into account the occupant's behavior habit changes. The advantage of such a system is to avoid discomfort zones for the occupant. This can be the case for instance when deactivating an HVAC system in a room predicted as unoccupied by a static model while occupants get in this room at such moments. The real-time error correction system proposed here will each time order to turn-on the HVAC system to heat or cool the room to meet the occupant's thermal comfort. Nevertheless, during the time of heating/cooling the room, there will be thermal discomfort for the occupant during all this period. In the case where the model is static (i.e. not adaptive), and if this kind of situation is repeated, the occupant may get too angry against the adoption of the system.

**Man in the loop:** or more precisely, occupant in the loop. In fact, all the work done, from the moment of finding that the building sector is the most energy consumers to the moment of predicting occupant behavior and optimizing energy, does not take into account the occupant interaction with the system itself. The occupant is that person who can help improve the system efficiently by customizing the notion of comfort. If he/she does not understand well the system or maybe feel insecure with such a new system, it can generate on him/her a high opposition to changes and thus jeopardize the benefits of all the previously cited solutions. Therefore, involving the occupant in these changes

by proposing a recommendation platform is a must. This allows him/her to understand the right habits to have, or send him/her notification when, for example, he/she forgets the windows open while the HVAC is turned-on in his/her room. Designing such an interaction and recommendation systems while putting the user experience (UX) at the heart of the approach is another important perspective for our work.

**Federated Learning-based privacy improvement:** in all our solutions, we use a centralized system, i.e. all our data are sent to one main server who analyzes and learns from this data, and takes a decision based on these data. Nevertheless, sending data to a server that may be in the cloud can represent a serious privacy issue. Fortunately, Federated Learning (or Collaborative Learning) trains algorithms across many decentralized edge devices (or servers). This helps hold the data locally without the need to transfer it and thus, minimize data privacy issues.





## List of Publications

- **Nour Haidar**, Nouredine Tamani, Yacine Ghamri-Doudane and Alain Bouju, "Occupant Behavior Prediction and Real-time Correction-based Smart Building Energy Optimization". IEEE Global Communications Conference (GLOBECOM), Taipei, Taiwan, 2020
- **Nour Haidar**, Nouredine Tamani, Yacine Ghamri-Doudane, and Alain Bouju. "Towards a new graph-based occupant behavior modeling in smart building". In 2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC).Tangier, Morocco, 2019.
- **Nour Haidar**, Nouredine Tamani, Felix Nienaber, Mark Thomas Wesseling, Alain Bouju and Yacine Ghamri-Doudane. "Data collection period and sensor selection method for smart building occupancy prediction". In 2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring), pages 1-6, Kuala Lumpur, Malaysia, 2019. Voir pages LIX et LXIV.





# Bibliography

- [1] United Nations. world population prospects 2019, 2019.  
see pages 14 and 18.
- [2] N. Haidar, N. Tamani, F. Nienaber, M. T. Wesseling, A. Bouju, and Y. Ghamri-Doudane. Data collection period and sensor selection method for smart building occupancy prediction. In *2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring)*, pages 1–6, 2019.  
see pages 15, 19, 88 and 95.
- [3] Nour Haidar, Nouredine Tamani, Yacine Ghamri-Doudane, and Alain Bouju. Towards a new graph-based occupant behavior modeling in smart building. In *2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)*, pages 1809–1814. IEEE, 2019.  
see pages 15, 19, 87 and 91.
- [4] Nour Haidar, Nouredine Tamani, Yacine Ghamri-Doudane, and Alain Bouju. Occupant behavior prediction and real-time correction-based smart building energy optimization. In *2020 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2020.  
see pages 16 and 19.
- [5] United Nations. Our growing population, 2020.  
see page 28.
- [6] R Newell, Daniel Raimi, and Gloria Aldana. Global energy outlook 2019: The next generation of energy. *Resources for the Future*, pages 8–19, 2019.  
see page 28.

- 
- [7] John Twidell and Tony Weir. *Renewable energy resources*. Routledge, 2015.  
see pages 28, 29 and 30.
- [8] Mehdi Ben Jebli, Slim Ben Youssef, and Ilhan Ozturk. Testing environmental kuznets curve hypothesis: The role of renewable and non-renewable energy consumption and trade in oecd countries. *Ecological Indicators*, 60:824–831, 2016.  
see pages 29 and 30.
- [9] U.S. Energy Information Administration (EIA). The united states uses a mix of energy sources, 2020.  
see page 29.
- [10] A Gürhan Kök, Kevin Shang, and Şafak Yücel. Investments in renewable and conventional energy: The role of operational flexibility. *Manufacturing & Service Operations Management*, 2020.  
see page 29.
- [11] Fortunat Joos, Gian-Kasper Plattner, Thomas F Stocker, Olivier Marchal, and Andreas Schmittner. Global warming and marine carbon cycle feedbacks on future atmospheric co<sub>2</sub>. *Science*, 284(5413):464–467, 1999.  
see page 29.
- [12] Anand Purushottam Pandit. Global warming-global climate change, evidences, causes and future effects. 2020.  
see page 29.
- [13] Habib Mohammad Naser, Osamu Nagata, Sarmin Sultana, and Ryusuke Hatano. Carbon sequestration and contribution of co<sub>2</sub>, ch<sub>4</sub> and n<sub>2</sub>o fluxes to global warming potential from paddy-fallow fields on mineral soil beneath peat in central hokkaido, japan. *Agriculture*, 10(1):6, 2020.  
see page 29.
- [14] GLOBAL ECOSOCIALIST. Oil spill threatens disaster for mauritius. 2020.  
see page 29.
- [15] Yuki Kudoh, Venkatachalam Anbumozhi, et al. Selecting the best mix of renewable and conventional energy sources for asian communities. *ERIA Research Project FY*, (26), 2015.

see page 30.

- [16] Aviel Verbruggen, Manfred Fischedick, William Moomaw, Tony Weir, Alain Nadaï, Lars J Nilsson, John Nyboer, and Jayant Sathaye. Renewable energy costs, potentials, barriers: Conceptual issues. *Energy policy*, 38(2):850–861, 2010.

see page 30.

- [17] Hoang Phong Le and Samuel Asumadu Sarkodie. Dynamic linkage between renewable and conventional energy use, environmental quality and economic growth: Evidence from emerging market and developing economies. *Energy Reports*, 6:965–973, 2020.

see page 30.

- [18] Andreas Coester, Marjan W Hofkes, and Elissaios Papyrakis. An optimal mix of conventional power systems in the presence of renewable energy: a new design for the german electricity market. *Energy Policy*, 116:312–322, 2018.

see page 30.

- [19] Xi Fang, Satyajayant Misra, Guoliang Xue, and Dejun Yang. Smart grid—the new and improved power grid: A survey. *IEEE communications surveys & tutorials*, 14(4):944–980, 2011.

see page 30.

- [20] Vehbi C Gungor, Dilan Sahin, Taskin Kocak, Salih Ergut, Concettina Buccella, Carlo Cecati, and Gerhard P Hancke. Smart grid technologies: Communication technologies and standards. *IEEE transactions on Industrial informatics*, 7(4):529–539, 2011.

see page 30.

- [21] Mohamed E El-Hawary. The smart grid—state-of-the-art and future trends. *Electric Power Components and Systems*, 42(3-4):239–250, 2014.

see page 30.

- [22] G Dileep. A survey on smart grid technologies and applications. *Renewable Energy*, 146:2589–2625, 2020.

see page 30.

- [23] James A Momoh. *Smart grid: fundamentals of design and analysis*, volume 63. John Wiley & Sons, 2012.  
see page 30.
- [24] Ryan Hledik. How green is the smart grid? *The Electricity Journal*, 22(3):29–41, 2009.  
see page 30.
- [25] Randal E Bryant, Randy H Katz, Chase Hensel, and Erwin P Gianchandani. From data to knowledge to action: Enabling the smart grid. *arXiv preprint arXiv:2008.00055*, 2020.  
see page 31.
- [26] John Theodore Houghton, Geoffrey J Jenkins, and Jim J Ephraums. Climate change: the ipcc scientific assessment. *American Scientist;(United States)*, 80(6), 1990.  
see page 31.
- [27] Vahid Vakiloroaya, Bijan Samali, Ahmad Fakhar, and Kambiz Pishghadam. A review of different strategies for hvac energy saving. *Energy conversion and management*, 77:738–754, 2014.  
see page 31.
- [28] International Energy Agency (2016). Tracking clean energy progress 2016. Retrieved from. Tracking clean energy progress 2016, 2016.  
see page 31.
- [29] Wen Wei Che, Chi Yan Tso, Li Sun, Danny YK Ip, Harry Lee, Christopher YH Chao, and Alexis KH Lau. Energy consumption, indoor thermal comfort and air quality in a commercial office with retrofitted heat, ventilation and air conditioning (hvac) system. *Energy and Buildings*, 201:202–215, 2019.  
see page 31.
- [30] World Meteorological Organization: WMO confirms 2019 as second hottest year on record. Wmo confirms 2019 as second hottest year on record, 01 2020.  
see page 31.

- [31] International Energy Agency (IEA) for the Global Alliance for Buildings and Construction (GlobalABC). 2019 global status report for buildings and construction towards a zero-emissions, efficient and resilient buildings and construction sector, 2019.  
see page 31.
- [32] Constantinos A Balaras, Elena G Dascalaki, Athina G Gaglia, Kaliopi Droutsas, and Simon Kontoyiannidis. Energy performance of european buildings. In *Energy Sustainability*, volume 47977, pages 387–396, 2007.  
see page 31.
- [33] Bart Poel, Gerelle van Cruchten, and Constantinos A Balaras. Energy performance assessment of existing dwellings. *Energy and Buildings*, 39(4):393–403, 2007.  
see page 31.
- [34] Hadi Shahnazari, Prashant Mhaskar, John M House, and Timothy I Salsbury. Heating, ventilation and air conditioning systems: Fault detection and isolation and safe parking. *Computers & Chemical Engineering*, 108:139–151, 2018.  
see page 31.
- [35] R Sendra-Arranz and A Gutiérrez. A long short-term memory artificial neural network to predict daily hvac consumption in buildings. *Energy and Buildings*, page 109952, 2020.  
see page 32.
- [36] Yibo Chen, Umberto Berardi, and Fengyi Zhang. Day-ahead prediction of hourly subentry energy consumption in the building sector using pattern recognition algorithms. *Energy*, page 118530, 2020.  
see page 32.
- [37] Daniele Antonucci, Ulrich Filippi Oberegger, Wilmer Pasut, and Andrea Gasparella. Building performance evaluation through a novel feature selection algorithm for automated arx model identification procedures. *Energy and Buildings*, 150:432–446, 2017.  
see page 32.



- [38] Chengliang Fan, Yundan Liao, and Yunfei Ding. Development of a cooling load prediction model for air-conditioning system control of office buildings. *International Journal of Low-Carbon Technologies*, 14(1):70–75, 2019.  
see page 32.
- [39] Riasat Sarwar, Heejin Cho, Sam J Cox, Pedro J Mago, and Rogelio Luck. Field validation study of a time and temperature indexed autoregressive with exogenous (arx) model for building thermal load prediction. *Energy*, 119:483–496, 2017.  
see page 32.
- [40] Hai Zhong, Jiajun Wang, Hongjie Jia, Yunfei Mu, and Shilei Lv. Vector field-based support vector regression for building energy consumption prediction. *Applied Energy*, 242:403–414, 2019.  
see page 32.
- [41] Chengliang Xu, Huanxin Chen, Weide Xun, Zhenxin Zhou, Tao Liu, Yuke Zeng, and Tanveer Ahmad. Modal decomposition based ensemble learning for ground source heat pump systems load forecasting. *Energy and Buildings*, 194:62–74, 2019.  
see page 32.
- [42] Sachin Kumar, Saibal K Pal, and Ram Pal Singh. A novel method based on extreme learning machine to predict heating and cooling load through design and structural attributes. *Energy and Buildings*, 176:275–286, 2018.  
see page 32.
- [43] Tao Liu, Chengliang Xu, Yabin Guo, and Huanxin Chen. A novel deep reinforcement learning based methodology for short-term hvac system energy consumption prediction. *International Journal of Refrigeration*, 107:39–51, 2019.  
see page 32.
- [44] Amin Mohammadi, Mohammad H Ahmadi, Mokhtar Bidi, Fatemeh Joda, Antonio Valero, and Sergio Uson. Exergy analysis of a combined cooling, heating and power system integrated with wind turbine and compressed air energy storage system. *Energy Conversion and Management*, 131:69–78, 2017.  
see page 32.
- [45] Huide Fu, Guiqiang Li, and Fubing Li. Performance comparison of photovoltaic/thermal solar water heating systems with direct-coupled photovoltaic

- pump, traditional pump and natural circulation. *Renewable Energy*, 136:463–472, 2019.
- see page 32.
- [46] Hong Li and Yue Sun. Operational performance study on a photovoltaic loop heat pipe/solar assisted heat pump water heating system. *Energy and Buildings*, 158:861–872, 2018.
- see page 32.
- [47] Bin Su. Impact of building envelope design on energy consumption of light structure school building. 2016.
- see page 32.
- [48] Chuang Wang, Da Yan, and Xiaoxin Ren. Modeling individual’s light switching behavior to understand lighting energy use of office building. *Energy Procedia*, 88(Supplement C):781–787, 2016.
- see page 32.
- [49] Patrícia Lourenço, Manuel Duarte Pinheiro, and Teresa Heitor. Light use patterns in portuguese school buildings: User comfort perception, behaviour and impacts on energy consumption. *Journal of Cleaner Production*, 228:990–1010, 2019.
- see page 32.
- [50] Lucia Ceccherini Nelli and Alberto Reatti. Smart active envelope solutions, integration of photovoltaic/thermal solar concentrator in the building façade. In *Green Buildings and Renewable Energy*, pages 459–467. Springer, 2020.
- see page 32.
- [51] Yonggang Zhang, Yongwei Zhong, Yingda Gong, and Lirong Zheng. The optimization of visual comfort and energy consumption induced by natural light based on pso. *Sustainability*, 11(1):49, 2019.
- see page 32.
- [52] Hamed Nabizadeh Rafsanjani and Ali Ghahramani. Towards utilizing internet of things (iot) devices for understanding individual occupants’ energy usage of personal and shared appliances in office buildings. *Journal of Building Engineering*, 27:100948, 2020.
- see page 33.

- [53] Oluwasikemi Ogunleye, Adeyemi Alabi, Sanjay Misra, Adewole Adewumi, Ravin Ahuja, and Robertas Damasevicius. Comparative study of the electrical energy consumption and cost for a residential building on fully ac loads vis-a-vis one on fully dc loads. In *Advances in Data Sciences, Security and Applications*, pages 395–405. Springer, 2020.  
see page 33.
- [54] Adeyemi Alabi, Oluwasikemi Ogunleye, Sanjay Misra, Olusola Abayomi-Alli, Ravin Ahuja, and Modupe Odusami. Comparative study of the electrical energy consumption and cost for a residential building with conventional appliances vis-a-vis one with energy-efficient appliances. In *Data, Engineering and Applications*, pages 297–311. Springer, 2019.  
see page 33.
- [55] Harsimranjit Kaur Romana and RK Malik. Energy consumption attributes in residential buildings-a case study of replacing conventional electrical appliances with energy efficient appliances. 2017.  
see page 33.
- [56] Paul Gill, Kate Stewart, Elizabeth Treasure, and Barbara Chadwick. Methods of data collection in qualitative research: interviews and focus groups. *British dental journal*, 204(6):291–295, 2008.  
see pages 33 and 34.
- [57] Edith D De Leeuw. Choosing the method of data collection. 2008.  
see page 33.
- [58] Margaret C Harrell and Melissa A Bradley. Data collection methods. semi-structured interviews and focus groups. Technical report, Rand National Defense Research Inst santa monica ca, 2009.  
see page 33.
- [59] Stephen L Schensul, Jean J Schensul, and Margaret Diane LeCompte. *Essential ethnographic methods: Observations, interviews, and questionnaires*, volume 2. Rowman Altamira, 1999.  
see page 33.

- [60] Andreas P Plageras, Kostas E Psannis, Christos Stergiou, Haoxiang Wang, and Brij B Gupta. Efficient iot-based sensor big data collection–processing and analysis in smart buildings. *Future Generation Computer Systems*, 82:349–357, 2018.  
see page 34.
- [61] Nguyen Cong Luong, Dinh Thai Hoang, Ping Wang, Dusit Niyato, Dong In Kim, and Zhu Han. Data collection and wireless communication in internet of things (iot) using economic analysis and pricing models: A survey. *IEEE Communications Surveys & Tutorials*, 18(4):2546–2590, 2016.  
see page 34.
- [62] Johnny KW Wong, Heng Li, and SW Wang. Intelligent building research: a review. *Automation in construction*, 14(1):143–159, 2005.  
see page 34.
- [63] Andrew Harrison, Eric Loe, and James Read. *Intelligent buildings in south East Asia*. Taylor & Francis, 2005.  
see page 34.
- [64] JA Powell. Intelligent design teams design intelligent buildings. *Habitat International*, 14(2-3):83–94, 1990.  
see page 34.
- [65] Diane J Cook and Sajal K Das. How smart are our environments? an updated look at the state of the art. *Pervasive and mobile computing*, 3(2):53–73, 2007.  
see page 34.
- [66] Diego Bonilla, Margarita Gil Samaniego, Rogelio Ramos, and Héctor Campbell. Practical and low-cost monitoring tool for building energy management systems using virtual instrumentation. *Sustainable cities and society*, 39:155–162, 2018.  
see page 35.
- [67] Patricia RS Jota, Valéria RB Silva, and Fábio G Jota. Building load management using cluster and statistical analyses. *International Journal of Electrical Power & Energy Systems*, 33(8):1498–1505, 2011.  
see page 36.

- [68] Bruna Faitão Balvedi, Enedir Ghisi, and Roberto Lamberts. A review of occupant behaviour in residential buildings. *Energy and Buildings*, 174:495–505, 2018.  
see page 37.
- [69] Elham Delzendeh, Song Wu, Angela Lee, and Ying Zhou. The impact of occupants’ behaviours on building energy analysis: A research review. *Renewable and Sustainable Energy Reviews*, 80:1061–1071, 2017.  
see pages 37 and 89.
- [70] Tianzhen Hong, Da Yan, Simona D’Oca, and Chien-fei Chen. Ten questions concerning occupant behavior in buildings: The big picture. *Building and Environment*, 114:518–530, 2017.  
see page 37.
- [71] Da Yan, William O’Brien, Tianzhen Hong, Xiaohang Feng, H Burak Gunay, Farhang Tahmasebi, and Ardeshir Mahdavi. Occupant behavior modeling for building performance simulation: Current state and future challenges. *Energy and Buildings*, 107:264–278, 2015.  
see page 37.
- [72] Tianzhen Hong, Sarah C Taylor-Lange, Simona D’Oca, Da Yan, and Stefano P Corgnati. Advances in research and applications of energy-related occupant behavior in buildings. *Energy and buildings*, 116:694–702, 2016.  
see page 37.
- [73] Shan Hu, Da Yan, Elie Azar, and Fei Guo. A systematic review of occupant behavior in building energy policy. *Building and Environment*, page 106807, 2020.  
see page 37.
- [74] Shady Jami, Zahra Sadat Zomorodian, Mohammad Tahsildoost, Maryam Khoshbakht, et al. The effect of occupant behaviors on energy retrofit: a case study of student dormitories in tehran. *Journal of Cleaner Production*, page 123556, 2020.  
see page 37.
- [75] Salvatore Carlucci, Marilena De Simone, Steven K Firth, Mikkel B Kjærgaard, Romana Markovic, Mohammad Saiedur Rahaman, Masab Khalid Annaqeeb, Silvia Biandrate, Anooshmita Das, Jakub Wladyslaw Dziedzic, et al. Modeling occupant behavior in buildings. *Building and Environment*, page 106768, 2020.

see page 37.

- [76] Isabella Gaetani, Pieter-Jan Hoes, and Jan LM Hensen. Estimating the influence of occupant behavior on building heating and cooling energy in one simulation run. *Applied Energy*, 223:159–171, 2018.

see page 37.

- [77] Mengda Jia, Ravi S Srinivasan, and Adeeba A Raheem. From occupancy to occupant behavior: An analytical survey of data acquisition technologies, modeling methodologies and simulation coupling mechanisms for building energy efficiency. *Renewable and Sustainable Energy Reviews*, 68:525–540, 2017.

see page 46.

- [78] Yixing Chen, Xin Liang, Tianzhen Hong, and Xuan Luo. Simulation and visualization of energy-related occupant behavior in office buildings. In *Building Simulation*, pages 1–14. Springer, 2017.

see pages 46 and 67.

- [79] Kaiyu Sun and Tianzhen Hong. A framework for quantifying the impact of occupant behavior on energy savings of energy conservation measures. *Energy and Buildings*, 146:383–396, 2017.

see pages 46 and 67.

- [80] Simona D’Oca, Stefano Corgnati, and Tianzhen Hong. Data mining of occupant behavior in office buildings. *Energy Procedia*, 2015.

see pages 46, 47 and 67.

- [81] Yohei Yamaguchi and Yoshiyuki Shimoda. A stochastic model to predict occupants’ activities at home for community-/urban-scale energy demand modelling. *Journal of Building Performance Simulation*, pages 1–17, 2017.

see pages 46 and 67.

- [82] Selin Yilmaz, Steven K Firth, and David Allinson. Occupant behaviour modelling in domestic buildings: the case of household electrical appliances. *Journal of Building Performance Simulation*, pages 1–19, 2017.

see pages 46, 47, 67 and 90.

- [83] Bing Dong, Zhaoxuan Li, and Gaelen Mcfadden. An investigation on energy-related occupancy behavior for low-income residential buildings. *Science and Technology for the Built Environment*, 21(6):892–901, 2015.  
see pages 46, 65 and 67.
- [84] Wei Wang, Jiayu Chen, and Tianzhen Hong. Modeling occupancy distribution in large spaces with multi-feature classification algorithm. *Building and Environment*, 2018.  
see pages 47 and 67.
- [85] Zheng Yang, Nan Li, Burcin Becerik-Gerber, and Michael Orosz. A systematic approach to occupancy modeling in ambient sensor-rich buildings. *Simulation*, 90(8):960–977, 2014.  
see pages 47, 67 and 89.
- [86] Luis M Candanedo, Véronique Feldheim, and Dominique Deramaix. A methodology based on hidden markov models for occupancy detection and a case study in a low energy residential building. *Energy and Buildings*, 148:327–341, 2017.  
see pages 47 and 67.
- [87] Luis M Candanedo and Véronique Feldheim. Accurate occupancy detection of an office room from light, temperature, humidity and co2 measurements using statistical learning models. *Energy and Buildings*, 112:28–39, 2016.  
see pages 47 and 67.
- [88] Zhaoxuan Li and Bing Dong. Short term predictions of occupancy in commercial buildings-performance analysis for stochastic models and machine learning approaches. *Energy and Buildings*, 158:268–281, 2018.  
see pages 47 and 67.
- [89] Anna A Adamopoulou, Athanasios M Tryferidis, and Dimitrios K Tzovaras. A context-aware method for building occupancy prediction. *Energy and Buildings*, 110:229–244, 2016.  
see pages 47, 67 and 90.
- [90] Zhaoxuan Li and Bing Dong. A new modeling approach for short-term prediction of occupancy in residential buildings. *Building and Environment*, 121:277–290, 2017.  
see pages 47 and 67.

- [91] Joaquim Massana, Carles Pous, Llorenç Burgas, Joaquim Melendez, and Joan Colomer. Short-term load forecasting for non-residential buildings contrasting artificial occupancy attributes. *Energy and Buildings*, 130:519–531, 2016.  
see pages 47 and 67.
- [92] John Hoffman. Self-directed work team transition: Leadership influence mediates self determination theory to describe variation in employee commitment. 2017.  
see page 47.
- [93] Irvan B Arief-Ang, Flora D Salim, and Margaret Hamilton. Da-hoc: semi-supervised domain adaptation for room occupancy prediction using co 2 sensor data. In *Proceedings of the 4th ACM International Conference on Systems for Energy-Efficient Built Environments*, page 1. ACM, 2017.  
see page 47.
- [94] Irvan Bastian Arief Ang, Flora Dilys Salim, and Margaret Hamilton. Human occupancy recognition with multivariate ambient sensors. In *Pervasive Computing and Communication Workshops (PerCom Workshops), 2016 IEEE International Conference on*, pages 1–6. IEEE, 2016.  
see page 47.
- [95] Chengdong Li, Zixiang Ding, Dongbin Zhao, Jianqiang Yi, and Guiqing Zhang. Building energy consumption prediction: An extreme deep learning approach. *Energies*, 10(10):1525, 2017.  
see page 66.
- [96] Anthony Tan. Occupancy estimation and tracking in indoor environment, 2020.  
see page 82.
- [97] J. Zhao, N. Frumkin, P. Ishwar, and J. Konrad. Cnn-based indoor occupant localization via active scene illumination. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 2636–2640, 2019.  
see page 82.
- [98] International Energy Agency iea. International energy outlook 2019 with projections to 2050, 2019.  
see page 88.



- [99] Microdata File. Commercial buildings energy consumption survey (cbeccs). 2015.  
see page 88.
- [100] Rania Elghamry and Neveen Azmy. Buildings orientation and it's impact on the energy consumption. 12 2017.  
see page 88.
- [101] Yu-Hao Lin, Kang-Ting Tsai, Min-Der Lin, and Ming-Der Yang. Design optimization of office building envelope configurations for energy conservation. *Applied Energy*, 171:336–346, 2016.  
see page 88.
- [102] Mengda Jia, Ravi S. Srinivasan, and Adeeba A. Raheem. From occupancy to occupant behavior: An analytical survey of data acquisition technologies, modeling methodologies and simulation coupling mechanisms for building energy efficiency. *Renewable and Sustainable Energy Reviews*, 68:525 – 540, 2017.  
see pages 88 and 89.
- [103] Tuan Anh Nguyen and Marco Aiello. Energy intelligent buildings based on user activity: A survey. *Energy and buildings*, 56:244–257, 2013.  
see page 88.
- [104] C Lee, J Tong, and V Cheng. Occupant behavior in building design and operation. In *HK joint Symposium 2014*, 2014.  
see page 88.
- [105] Mehdi Maasoumy and Alberto Sangiovanni-Vincentelli. Smart connected buildings design automation: Foundations and trends. *Foundations and Trends® in Electronic Design Automation*, 10:1–143, 03 2016.  
see page 89.
- [106] Frédéric Haldi, Davide Calì, Rune Korsholm Andersen, Mark Wesseling, and Dirk Müller. Modelling diversity in building occupant behaviour: a novel statistical approach. *Journal of Building Performance Simulation*, 10(5-6):527–544, 2017.  
see page 89.

- 
- [107] Jakub Wladyslaw Dziedzic, Da Yan, Hongsan Sun, and Vojislav Novakovic. Building occupant transient agent-based model–movement module. *Applied Energy*, 261:114417, 2020.  
see page 89.
- [108] Andrew Kusiak, Mingyang Li, and Fan Tang. Modeling and optimization of hvac energy consumption. *Applied Energy*, 87(10):3092–3102, 2010.  
see page 89.
- [109] Jonathan Brooks, Saket Kumar, Siddharth Goyal, Rahul Subramany, and Prabir Barooah. Energy-efficient control of under-actuated hvac zones in commercial buildings. *Energy and Buildings*, 93:160–168, 2015.  
see page 89.
- [110] Zhenghua Chen, Jinming Xu, and Yeng Chai Soh. Modeling regular occupancy in commercial buildings using stochastic models. *Energy and Buildings*, 103:216–223, 2015.  
see page 89.
- [111] Chenda Liao, Yashen Lin, and Prabir Barooah. Agent-based and graphical modelling of building occupancy. *Journal of Building Performance Simulation*, 5(1):5–25, 2012.  
see page 90.
- [112] John G Kemeny and J Laurie Snell. *Markov chains*. Springer-Verlag, New York, 1976.  
see page 103.
- [113] Robert E. Schapire. *Explaining AdaBoost*, pages 37–52. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.  
see page 103.

