# User Profiling in the Intelligent Office

SAIFULLIZAM PUTEH

A thesis submitted in partial fulfilment of the requirements of

Nottingham Trent University for the degree of

*Doctor of Philosophy*

September 2013

This thesis is dedicated to my parents, wife and family with great gratitude. Undoubtedly, without their prayers and support this thesis would have been impossible.

# Acknowledgements

The work described in this thesis was carried out at Nottingham Trent University, between July 2009 and September 2013; while the researcher was working as a full-time doctoral research student.

This thesis could not have been accomplished without the help of three outstanding individuals, namely Dr Ahmad Lotfi, Dr Caroline Langesiepen, and Dr Mohammad Javad Akhlaghinia. First of all, I owe a great deal of gratitude to my director of study, Dr Lotfi, for his continual support, in-depth knowledge, and kindness. Dr Lotfi is an extremely busy man, but was always able to find time to talk, help, and encourage me; and for that, I am forever in his debt.

I would like to thank my second supervisor, Dr Lagensiepen, for nurturing excellent ideas within these research areas and for believing in me. As an extremely busy woman, she was always ready to offer words of encouragement and impart sound and valuable advice.

I would also like to thank Dr Akhlaghinia for sharing his research ideas and experiences, and making me confident enough to deal with developing my data collection system for this project.

For my three supervisors, for whom I have the utmost respect, I thank you all for your continued support and kindness.

Actually, there are a number of parties that need to be thanked for their continued support and words of encourage throughout this PhD.

I would like to thank the School of Science and Technology, Nottingham Trent University, for providing me with excellent research activities and workshops during my research, and support from the Ministry of Higher Education, Malaysia, and Tun Hussein Onn, University of Malaysia.

I would like to thank all of my NTU friends, for when things did not go to plan; they were there for me in more ways than one, and we shared those experiences together. There are other kind parties that I would also like to thank, such as the Nottingham Malaysian Community, for their family and social support, after my family returned to Malaysia from Nottingham, and I lived alone.

For my parents and family, I cannot put into words the kindness, support, and love that you have shown to me. I thank you all from the bottom of my heart. God bless you all.

**Saifullizam Puteh**

**September 2013**

# Abstract

The research aim is to investigate different methods of profiling user activities in an office environment. This will allow optimal use of resources in future Intelligent Office Environments while still taking account of user preferences and comfort. To achieve the goal of this research, a data collection system is designed and built. This required a wireless Sensor Network to monitor a wide range of ambient conditions and user activities, and a software agent to monitor user's Personal Computer activities. Collected data from different users are gathered into a central database and converted into a meaningful format for description of the worker's Activity of Daily Working (ADW) and office environment conditions.

Different techniques including Approximate Entropy (ApEn), consistency measures, linear similarity measures and Dynamic Time Warping (DTW) are employed to quantify a user's behaviour and extract a user profile. The individual user profile is representative of a user's preferences, consisting of user routine activities, consistency of office usage and their thermal comfort. Using the statistical techniques, consistency and ApEn, it is possible to characterise different users with only a few parameters. Using similarity techniques one can assess the interrelationship of different aspects of a user's behaviour.

This helps to assess the importance of those aspects within the profile. The novel contribution is the use of these techniques within the context of ADW.

This research investigates soft computing techniques to enhance user profiling. A novel fuzzy characteristic matrix is proposed to summarised the ADW. The activity recognition models using an event-driven and a fuzzy inference system are proposed to recognise a worker's activities during times when the office is occupied and unoccupied during a workday. The experimental results demonstrate the models recognise a worker's activities and can classify into six categories (home, lunch, short break, out of office duties, not use computer/lighting and use computer/lighting) with accuracy of more than 90%.

# Publications

The following publications have been published as a direct result of this thesis:

**Refereed Conference Papers**

- S. Puteh, A. Lotfi and C. Langensiepen, "Activities Recognition in Intelligent Office Environment", Proc. of Smart Offices and Other Workplaces workshop. Athens, July 2013.

- S. Puteh, C. Langensiepen and A. Lotfi, "Fuzzy Ambient Intelligence for Intelligent Office Environments", WCCI 2012 IEEE World Congress on Computational Intelligence (FUZZ-IEEE), June, 10-15, 2012 - Brisbane, Australia, pp. 41-46.

- S. Puteh, C. Langensiepen, A. Lotfi, "Similarity Pattern Mining in Intelligent Office Environments", in Workshop Proc. of the 7th Int. Conference on Intelligent Environments (IE' 11), Nottingham, UK, 25-28 July, 2011, pp. 562-573.

- S. Puteh, C. Langensiepen, A. Lotfi, "University Office Simulator for Energy and Comfort Optimisation", in Proc. of the 25th European Conference on Modelling and Simulation (ECMS 2011), Krakow, Poland, 7-10 June, 2011, pp. 430-435.

# Contents

# List of Figures

# List of Tables

# Nomenclature

**Roman Symbols**

ADL Activities of Daily Living

ADW Activities of Daily Working

ANN Artificial Neural Network

BMS Building Management System

CI Computational Intelligence

DTW Dynamic Time Warping

FIS Fuzzy Inference System

FSM Finite-State Machines

GL Gower and Legendre Similarity Measure

HD Hamming Distance

JCD Jaccard Similarity Measure

PC Personal Computers

PCA Principal Component Analysis

PIR Passive Infra-red

PMV Predicted Mean Vote

PPD Predicted Percentage of Dissatisfied

RFID Radio Frequency Identification

WSN Wireless Sensor Network

# Chapter 1

# Introduction

Buildings are becoming the fastest growing energy consuming sector. Applying energy efficiency measures could contribute to the reduction of current energy consumption. To be able to apply energy efficiency measures, it is required to interact with the environment. The availability of modestly priced sensors and low cost computers allow us to consider individualised monitoring and control of the environment.

In this research we are specifically investigating ways to improve the energy efficiency in an office environment. Apart from the energy optimisation issue, there are other factors such as office worker's performance which could be investigated. As reported in [1], one of the major causes of stress for clerical workers is the lack of control of their environmental conditions. If the energy consumption units including lighting, heating and Personal Computer (PC) are made more responsive to the user's habits, routines and preferences, there would be more acceptance of their use.

An office environment equipped with appropriate sensory devices and actuators is required to be able to control the environmental conditions. Such an environment will be referred to as “Intelligent Office” environment. Apart from the monitoring and control of the environment, there should be an intelligent decision-making process taking into account the office user’s work activities and personal preferences. This will be referred to as a Building Management System (BMS). Studies in [2–5] have highlighted the importance of different sensors and a wide range of added capabilities in BMS including building security, activity recognition and automation control system.

The rest of this chapter is structured as follows: in the next section an overview of this research is presented. In Section 1.2, the aim of this thesis and the proposed objectives are presented. Section 1.3 introduces the major contribution of the thesis. Finally, the remaining chapters of this thesis are outlined in Section 1.4.

## 1.1 Overview of the Research

In modern office environments, lighting systems, heating/cooling system and PC are the main energy consumers. Many companies would like to reduce their energy usage for two reasons. The first is that they worry about the environment and want to reduce the impact they have on it. The second reason, and most likely the one that companies care about most, is cost. For example, PCs waste a lot of energy due to being left on for long periods of time when not in use. Even though they have power management modes to reduce their energy consumptions when they are not in use, these are not always being used.

Some workplaces incorporate reactive systems such as Passive Infra-red (PIR) activated lighting, but these can be activated/deactivated inappropriately. Heating systems often work on the assumption of a 9:00 AM to 5:00 PM presence, five days a week, whereas an individual office worker may have a different schedule, including long periods out of the office. Similarly, automated office computer shut down may be set based on assumptions of behaviour that are inappropriate, leading users to try to find ways of subverting the mechanisms so that their computer remains on and avoids the inconvenience of a slow restart.

The proposed research framework is illustrated in Figure 1.1. There are three distinct phases to develop the research; in the first stage, a data collection system is developed. Different monitoring and data collection system are investigated. The data collection system collects environmental conditions, user activities and office conditions. In the second phase, data mining techniques are applied to identify different user characteristics. Similarity measures are used to compare users' behaviour, detect similar behaviour between different users and also compare the user's behaviour across different days/weeks. In the third stage, activities recognised from a user are represented as user profile. User profile is used to summarise the activities of a user in an office environment. User profile is also used to optimisation environmental control system, so that the office condition are adjusted in line with the individual user.

## 1.2 Aims and Objectives

The research question in this project is how to understand and characterise the behaviour of an office user in terms that can be used for optimising comfort? To answer the research question, the following aim is identified.

The aim of the research is to record and analyse the detailed behaviour of users

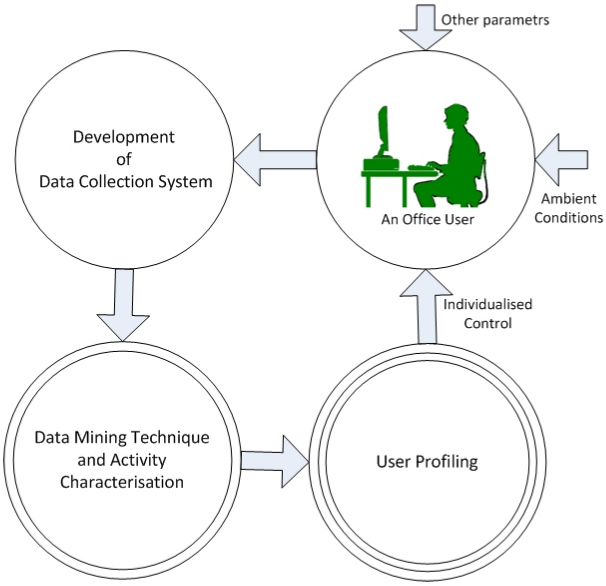



Figure 1.1: Research Framework

in office environments. Activities are monitored using low level sensory devices to detect when office workers enter/leave the room, sit down at the desk, and use the PC. We also monitor when they switch the light on/off, adjust the heating or leave the room to get a drink. Using this low level information gathered from the office environment, a user profile will be created such that the environment could be controlled based on the learned profile.

In order to accomplish the aim of this research, the following objectives are identified:

1. To investigate the requirements for the development of an appropriate monitoring system for intelligent office environment.

2. To investigate the suitability of a Wireless Sensor Network (WSN) to measure all relevant parameters with minimum interference with the users' daily activities.

3. To integrate collected data from different sources and present them in a uniform format. Collected data from different offices must be in a central database and it is essential to store the information in a suitable format for further processing.

4. To investigate efficient and meaningful formats of representation and visualisation of raw data. Raw data in either binary or analog format may not representing the activities in a meaningful format and it is important to convert the raw data into another format to make it more understandable.

5. To extract a user profile based on the collected data from an office worker. The profile represents the user activities in a simple and more meaningful format. User routine activities, consistency of office usage and also user ambient condition preferences should be represented within the user profile.

6. To investigate soft computing techniques to enhance user profiling and create more specific user characteristics which will be use to identify different activities. The user characteristics will form the basis of controlling the lighting, heating and the PC.

## 1.3 Major Contributions of the Thesis

The main contributions of this thesis are:

- Development of software and hardware required to collect ambient conditions, user activities and office conditions in intelligent office environments. Both wired and wireless systems are investigated for their suitability in an office environment.

- Collect data that represent Activities of Daily Working (ADW) for different office users. The ADW are related to university academic offices representing academic staff working activities, which involve more varied patterns of behaviour.

- Investigate and determine similarities between different users' ADW and also compare the users behaviour across different days/weeks. Both linear and non-linear similarity measures are applied. Different techniques including Dynamic Time Warping (DTW) are investigated. The novel contribution is the use of these techniques within the context of ADW.

- Using statistical techniques to quantify users' behaviour and their preferences and produce a novel signature representing the individual's ADW.

- Proposed a novel fuzzy characteristic matrix to summarised the activities of an office user. The characteristic matrix is presented as fuzzy values which indicate the likelihood of the sequence of activities.

## 1.4 Thesis Outline

This thesis consists of seven chapters that are summarised as follows:

**Chapter** 2: Literature Review

This chapter gives a review of the relevant literature related to different aspects of intelligent offices. Initially, an introduction to energy and comfort performance of building is presented followed by outline of building monitoring technologies, building occupant behaviour, human activities recognition and survey of intelligent building environments.

**Chapter 3**: Experimental Architectures

This chapter reports the experimental architecture development to monitor the ADW in an office environment. The system's development and structure are explained in detail including the WSN for measurement of ambient conditions, the PC monitoring application, the database where all activities are logged and the control server application. A simulator is also developed to generate equivalent pattern to a office worker. The simulator will allow more repeatable testing and assessment of algorithms developed in other chapters.

**Chapter 4**: Data Analysis Techniques

Both statistical and similarity measure techniques used to generate simple user profile are presented in this chapter. A brief review of linear similarity measures and the DTW algorithm (non-linear similarity measure) used to measure the differences between user patterns are presented in this chapter. Application of the techniques presented in this chapter are presented in Chapter 5.

**Chapter 5**: User Profiling

This chapter explains the analysis of the office worker's behaviour in an office environment. Statistical techniques namely consistency measures and approximate entropy measures are used to construct an individual user profile for energy and comfort optimisation. In this chapter, similarity measure techniques are also used to identify similarities between user activities. Some experimental results are presented demonstrating the similarities and dissimilarities in user activities with respect to the energy usage. Different techniques including hierarchical clustering and Principal Component Analysis (PCA) are used to identify energy usage behaviour.

**Chapter** 6: Enhance User Profiling

In this chapter user characteristics are used to summarise user activities. A fuzzy characteristic matrix is proposed to represent the user activities based on start-time and duration of an activity. Sensor data gathered from users are used to create these models. Then, the chapter introduces an event-driven model and a fuzzy inference system for recognising user activities. In an event-driven model, IF-THEN rules are used to determine the transition events. The fuzzy rules and the membership functions of activity recognition are defined based on sensory signals and temporal variables. The experimental results are presented, the proposed activity recognition methods have demonstrated able to identify ADW of a user, either she/he in office and out of office.

**Chapter** 7: Conclusions and Future Works

This chapter provides the pertinent conclusions arise from this thesis and formulates some future research in monitoring the daily activations of the worker in office environments for office energy and comfort optimisation.

# Chapter 2

# Literature Review

## 2.1 Introduction

Nowadays, the energy performance of building and occupant comfort are both priorities. The aim of building technology is to achieve energy efficiency and at the same time increase occupants' satisfaction. Rapid progress in software and hardware have made it possible to improve the intelligence of BMS. This is necessary in an era of increasing energy costs. In order to help reduce all these problems, the "intelligent building" industry need to be further explored. Intelligent building system should allow integration and automation of all technologies and computational intelligence techniques to optimise energy consumption [6], occupants' well-being [7], safety [8] and work productivity [9].

Figure 2.1 shows that energy consumption in the United Kingdom has continued to increase since 1970. Although, buildings in domestic and services sectors have had less attention regarding production of Carbon Dioxide ($CO_2$) emission than vehicles and industrial machines, improvement in building performance to

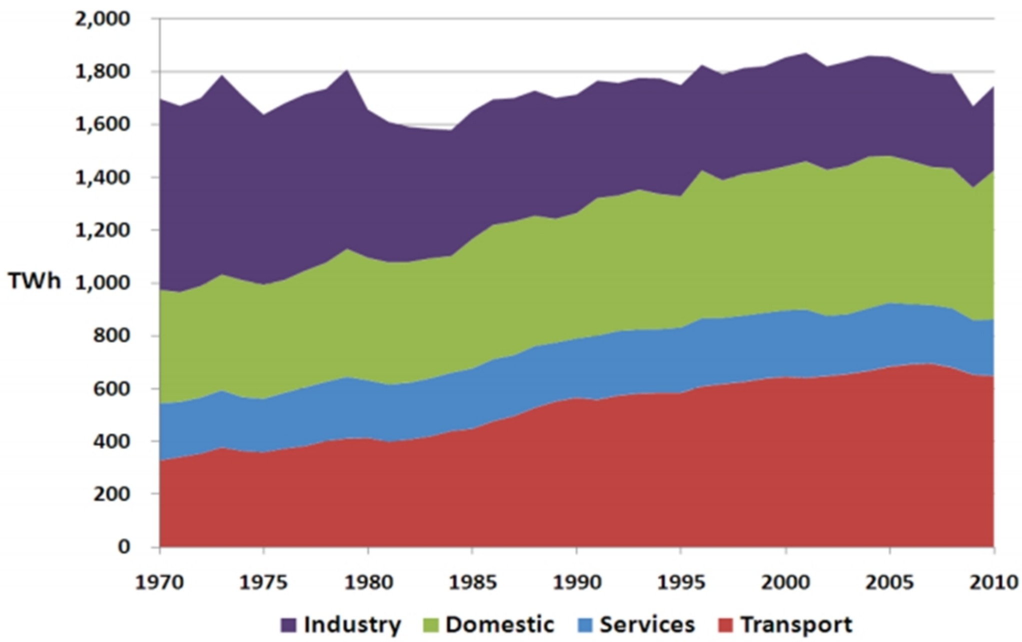



Figure 2.1: Energy consumption in the UK by sectors [10].

achieve energy efficiency and maintain thermal comfort [11], would also help to reduce $CO_2$ emissions.

Energy consumption by energy source and sector in the UK is shown in Figure 2.2. Electricity is the main energy source followed by gas for the buildings sector. Thus electricity management is a big part of BMS. Improving the effectiveness of BMS will not only benefit the building occupants, but help long term energy sustainability [12].

## 2.2 Energy Efficiency in Buildings

The energy cost of a building is directly proportional to its capacity and hours of energy usage. For example, the study in [13] used private office buildings

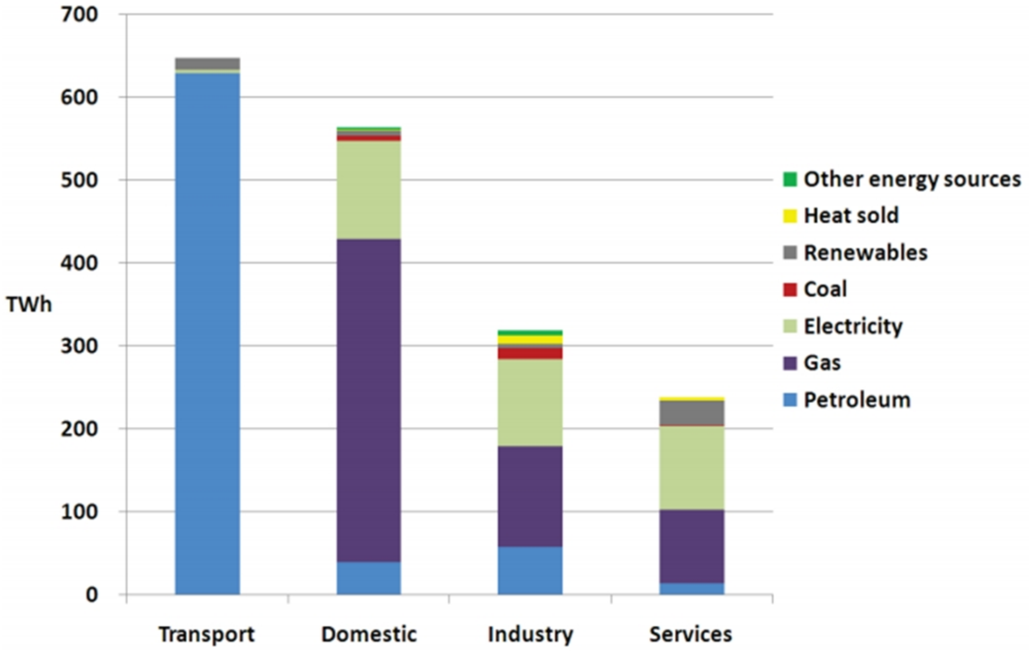



Figure 2.2: Energy consumption by energy source and sector in the UK (2010) [10].

to find the relationship between energy usage and operating hours. In another study [14],it is found to be difficult to optimise the thermal comfort preferences of individuals in the office at all times. The important factors to be considered is that each individual has different thermal comfort satisfaction, and at different times and places an individual has different preferences of comfort in an office environment. In [15] they improved office efficiency based on heating, cooling and lighting to propose a future office environment. The study in [16] showed that social and personal factors can influence one's perceived health and comfort. In order to investigate individual comfort, it is necessary to investigate the relationship between personal, social and building factors.

Looking into the future of residential and office buildings, Mitsubishi Electric Research Labs (MERL) has collected motion sensor data from a network of over

200 sensors for two years in a 2-floor office environment [17]. This data is the residual trace from the people working in MERL. The dataset has been made publicly available as a benchmark to identify social and individual behaviour in office environment. Similar datasets have been collected in a home environment as part of CASAS smart home project [18].

Many issues apply to BMS, such as how to design and operate to comply with standards [19], sustainability [20], and maintaining comfort for everyone [21]. Research work needs to focus on designing, inventing and manufacturing intelligent building technology. At the education and society level, energy-awareness campaigns need to encourage people to save energy, whilst human resource programmes need to produce competence people to retrofit, operate and maintain in a manner that reduces the use of energy.

### 2.2.1 Building Characteristics

The largest portions of energy consumption by consumers are space heating, cooling and office equipment [22]. In addition, [23] reported that residential homes, commercial sector, offices, warehouses and premise are the main contributors to energy consumption and carbon emissions. According to these reports, we can state that office buildings are an important sectors to focus on, in order to achieve energy efficiency and at the same time improve indoor thermal comfort.

Kazanasmaz et al. [24] developed an office building prediction model to determine daylight luminance using Artificial Neural Networks (ANNs) . A study in [25] has carried out research on the daylight pattern depending on movement of sun, latitude of the building, climate condition, ambient temperature and sunshine availability. Although their work is aimed at designs to help predict luminance within buildings, the predictions could also be used once the building is in use to ensure that the artificial lighting is switched on/off before the occupants realised that they needed it [26].

### 2.2.2 Energy and Comfort Monitoring

Different occupant satisfaction levels and weather influence thermal comfort and the perception of thermal conditions. The study in [2] has developed a Multi-agent Control System to improve the efficiency of control systems for indoor environments including user preferences. This study took into consideration users' preferences on thermal, luminance comfort, indoor air quality and energy conservation.

Due to the differences in internal heat load, the characteristic differences between zones over the building and individual physiological/psychological differences, it is impossible to satisfy everyone with the same indoor condition provided to all occupants [14]. Investigation into the comfort levels of pupils by [19, 27, 28] showed that the wrong temperature in the classroom led to poorer learning performance. Similarly, performance studies of call centre workers in Sacramento [29] showed that light levels, ventilation status and temperature all had significant effects on performance, though the effects are intertwined and complex. Seppanen et al. [30] have collated studies on temperature and performance, and provide convincing evidence of the importance of maintaining the office temperature between 21-25°$C$ to optimise performance. Aries et al [31], for example, use multiple survey items to assess worker discomfort, sleep quality and hindrance, in order to relate building aspects to any physical and/or psychological effects. However, Haynes [9] points out that while there is sufficient evidence to support claims that office comfort affects productivity, there is no agreement as to how office comfort should be measured.

The workplace research by Knoll [32] stated that enhancing environmental control can improve users' performance and productivity. Rashidi [33] highlights the fact that it is important to monitor and recognize all activities that the worker regularly performs in their working environment.

Thermal comfort can be affected by heat transfer such as conduction, convection, radiation, and evaporation heat loss. There have been a number of studies relating thermal comfort to human psychology factors [34]. Behavioural adaptation such as the occupant's clothing, taking hot/cold drinks, etc. are psychological adaptations and affect the capability to adapt to the thermal environment [35]. Fanger's formulas [36] is based on average criteria for population comfort and it is widely used in themal standards such as Predicted Mean Vote (PMV) and Predicted Percentage of Dissatisfied (PPD) [37, 38].

Canadian Centre for Occupational health and Safety (CCOHS) [39] have suggested the thermal comfort setting, as shown in Table 2.1. This table is an adaptation from American Society of Heating, Refrigerating and Air Conditioning Engineers (ASHRAE) standard [40]. This table is based on numerous comfort evaluations under controlled steady state conditions using thousands of randomly selected respondents [41].

## 2.3 Intelligent Buildings

The rapid development of intelligent technology such as sensors, computer technology, simulation technology and system networking have provided opportunities to generate a high degree of intelligence in environment control systems. For example, a study in [42] designed an intelligent residential lighting control system based on a ZigBee wireless sensor network and fuzzy controller. Experimental results showed that the system contributed to economy and energy efficiency. In [43–45] it is shown that simulation tools for thermal comfort are capable of quantifying the relationship of surface materials, indoor and outdoor environments. Focusing on the intelligent building, a study in [46] developed eight key intelligent building indicators in order to construct models for appraising intelligent building systems.

In [47, 48], surveys are conducted to assess the weighting given to intelligent criteria of buildings. The results showed that work efficiency is the most important selection criterion for various intelligent building systems, while user comfort, safety and cost effectiveness are also considered to be significant. This finding indicated that any intelligent building system must perform efficiently to contribute to practicality and building occupants' satisfaction. Research by [49],

Table 2.1: Ranges of temperature and relative humidity for offices (adapted from ASHRAE standard).

| Temperature and Humidity Ranges for Comfort | | |
|---|---|---|
| Condition | Relative Humidity | Acceptable Temp. ($°C$) |
| Summer(light clothing) | If 30%, then | 24.5 - 28 |
| Summer(light clothing) | If 60%, then | 23 - 25.5 |
| Winter (warm clothing) | If 30%, then | 20.5 - 25.5 |
| Winter (warm clothing) | If 60%, then | 20 - 24 |

took into account the BMS to monitor daily energy operations in order to support the decision making process of selecting energy saving measures.

A BMS is needed to integrate all intelligent building aspects to contribute comprehensive strategic management, in order to analyse and report the building performance, and then provide data analysis to any decision making system. It aims to provide intelligent functionality to respond to the energy demand and comfort of building environments for normal daily operation [49].

## 2.4 Human Behaviour Recognition

Human activities recognition in buildings is the subject of interest for many researchers. Activities recognition in home environments [33, 50–56], hospital environment [57, 58] and the office environment [59–61] are being investigated.

Liming et al. in [5] have conducted research based on sensor-based human activity recognition. They suggested a complex process of activity recognition that can be approximately classified by four basic steps. They are:

1. to choose and deploy appropriate sensors in order to monitor and capture user's behaviour in building.
2. to collect, store, and process perceived information for data representation at appropriate level.
3. information gathered from monitoring place based on user's activity daily living are used to create computational activity models.
4. to select and develop intelligent algorithms to infer activities from sensory data.

Human activity recognition and pattern discovery is explained in [62]. Tabak and deVries [63] examined the intermittent activities that interrupt the planned "normal" activities of office workers. They found that probabilistic and S-curve methods could be used to predict activities (such as "smoking", "go to toilet") and these could be used in fine grained simulations of building performance. However, they cautioned that the results applied to typical Dutch office based organisations, and other office environments might need further experiments to generate data.

More survey results based on the human activities monitoring system and activities recognition techniques are presented in the following sections.

### 2.4.1 Activities Monitoring

People's activity in a building is based on their schedule of work, lifestyle and social activity. Xin et al. [64] mentioned that one of the key features of an intelligent environment is to provide monitoring Activity of Daily Living (ADL) . In many studies [50,55,58,64–67], ADL are monitored to assess elderly people's activity in the home environment, and attempts are made to process activity sequences to make them more understandable. For example, daily home activity involves basic functions like preparing breakfast or food, showering, walking, sleeping, watching television, reading books etc.

Recently, advanced intelligent sensor technology has resulted in various types of sensors that have been used by researchers to get features from activity monitoring. A study in [54] has used physiological sensors (cardiac frequency, activity or agitation, posture and fall detection sensor), microphones, PIR sensors, door sensors and state-change sensors. This monitoring system has been combined with a fuzzy logic system for recognizing activities in readiness for the next generation of smart houses. David Naranjo et al. [68] have proposed hardware and software design and implementation of low-cost, wearable, and unobtrusive intelligent sensors for monitoring human physical activities. Many researchers have successfully conducted research with similar ideas in activity monitoring systems, such as Interactive Continuous Autonomic Logging and Monitoring (iCALM) [69], human activity recognition in pervasive health-care systems [58] and occupancy monitoring system [70].

### 2.4.2 Activities Recognition Techniques

To recognise human activities, different computational techniques have been applied. Data mining techniques including Discontinuous Varied-Order Sequential Miner (DVSM) [33], classification tree methods [71–74] and Hidden Markov Models(HMMs) [75–77] have been used. More intelligent computational techniques including Fuzzy Inference Systems(FIS) in activity detection [78, 79] and some hybrid soft computing approaches namely neuro-fuzzy techniques [80, 81] are investigated. Some statistical methods and Bayesian classifier are discussed in [74, 82, 83].

As mentioned previously, when sensor information is collected and activities are detected based on individual activity, the intelligent environment needs a model to track each activity and also to recognise forthcoming activity to help people improve their quality of life [33]. Song et al. mentioned in [84] that behavioural pattern mining is a significant process to recognise the relationship and limits between real user's behaviour and an event log. Therefore, the model of behaviour pattern mining needs a process to determine the form of the mined results. Consequently, the process model influences the design of the mining algorithm and the related approach. They introduced the major stages of behaviour pattern mining as summarised below:

1. **Events recording:** This stage arranges the original data for behaviour pattern mining. The basic types of information include activity, event, participant, and time. The logged data may be extended as appropriate for any particular requirement.

2. **Prepossessing:** This stage provided the log format to fit the mining method, reduces noise and checks for the record. As a characteristic of behaviour pattern mining, common event logs may be passed on to make use of existing work-flow mining methods.

3. **Mining and discovering:** Implement mining algorithm to discover the behaviour patterns.

4. **Verification:** This stage applies the discovered results to system improvement and research, and checks out if the results fit the experience of the real world.

A study in [85] used a semantic-based similarity to update user profiles representing user preferences. They have monitored internet usage of a user to create a profile for each user. Based on the user profile, they have provided relevant information to the internet user. Ghulam et al. [86] proposed an agent-based user profiling to analyse user behaviour on the web for constructing user profile. They categorised the user profile into demography and user interest. Iglesias et al. [87] used a classifier to create a user behaviour profile. The classification method is used to develop user profiles based on computer activity.

Danni Wang et al. [88] used statistical properties of occupancy in single person office. Their work proposed a probabilistic model to predict and simulate occupancy in a single person office. A study in [89] proposed a Thermally Activated Building System (TABS). Their experimental work has shown that TABS model demonstrated that it is able to cope with user behaviour such as time of arrival, departure and temporary absence. Taherian et al. [90] also contribute in this area. However, their work identifies the user behaviour based purely on electricity usage in households and office spaces to produce an energy profile.

## 2.5 Data Representation

In order to store the collected data from the environment, it is essential to represent them in an appropriate format. Data will be collected from different sources with different time scale and even different format (analog and binary). To investigate the best form of data representation, many researches have been conducted in this area. The following sections will review these.

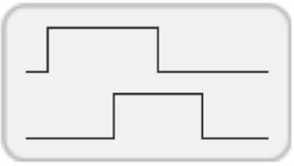
### 2.5.1 Temporal Data Representation

Lee et al. [91] have been discovering knowledge from temporal interval data in their research. According to Allen's theory about an algebra of temporal relations on interval [92], they proposed a new temporal data mining technique to extract temporal interval relation rules from temporal interval data. Lee et al. claimed that the proposed approach increases the model efficiency to identify activity from temporal interval data. Nazerfard et al. [53] proposed a framework for Discovering of Temporal Features and Relation of Activities (DTFRA). This framework discovers activity from data based on start time and duration using k-means clustering. The proposed approaches by Lee et al. and the DTFRA framework have shown that both approaches are able to predict start time intervals and represent the temporal features of activity patterns [53, 91].

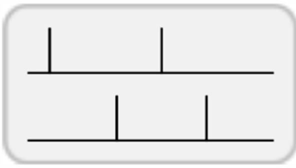
### 2.5.2 Time Series Representation

Time series analysis is often used in multi-sensory data representation. Kasteren [93] has discussed the advantages and disadvantages of time series with discrete time intervals for the recognition process. He determined that if an interval is very small, it would incorporate signal noise, while a moderately large interval will smooth out important aspects of the signal. He proposed different feature representations, i.e. change the binary data $X(t)$ to change points $Xt$ as shown in Figure 2.3. Binary data representation uses the sensor data directly and '1' indicates that the sensor is activated and '0' is inactivated. The change point representation indicates when a sensor event takes place, i.e. showing '1' when it changes state.

Akhlaghinia et al. in [70, 94], proposed a signal technique to combine time series of sensory signals in different areas for home occupancy monitoring. They used multiple PIR sensors to monitor single occupant activities in a home environment. The conditions for this approach are that there is no parallel activity in the different areas detected, and that all activities are genuinely from a single

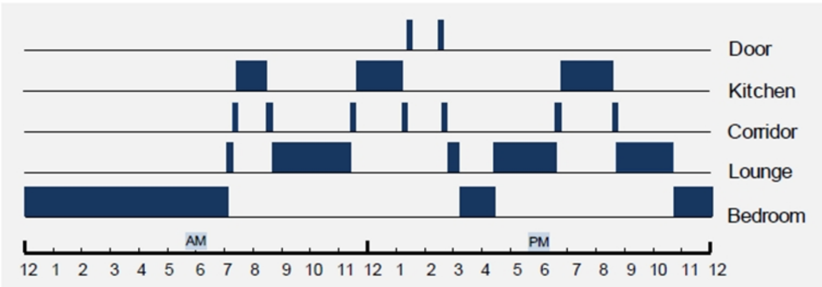


(a) Binary data

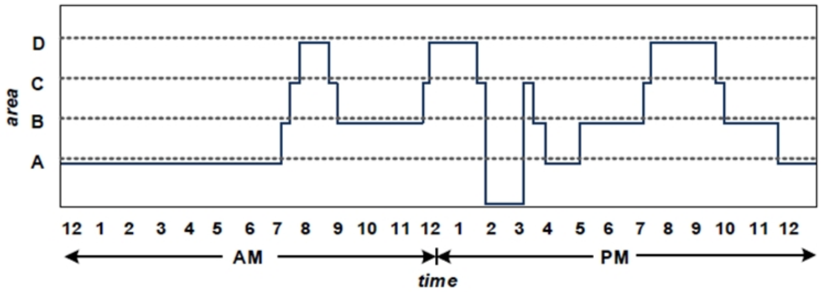


(b) Discrete-time signal

Figure 2.3: Different time-series feature representation [93]: a) binary data representation b) change point representation.




(a) Individual PIR Signals




(b) Combined Time-Series

Figure 2.4: Time-series representation of PIR sensors [70]: a) activities in different areas b) combined signals.

user. The occupancy signals as shown in Figure 2.4-a can be transformed into a combined occupancy time-series shown in Figure 2.4-b [70]. In Figure 2.4(b), each level represents the occupancy of a certain area.

### 2.5.3 Gray Coded Binary Representation

Many researchers have applied binary sequence data based on sensory signal in their pattern recognition process [70, 93, 94]. However, there are many issues still present and we need new exploration mechanisms in order to optimise the activity recognition model. For example, Akhlaghinia et al. in [70, 94] have proposed a combination signal technique to combine binary signal for PIR sensors. However, the proposed approach did not allow for overlapping. Therefore, when recording parallel activities from an observation area, binary representation is the best way to record all information. However, it is not as easy to combine all the signals based on bit order or using significant bit position for conversion to decimal or hexadecimal, because it is will increase the distance between one code word and the next [95]. Gray code is associated with many fields such as mechanical position sensors, electronic circuit, and so on [95, 96].

## 2.6 Office Building Monitoring Technologies

The main factors driving better building operation are the provision of a healthy environment, thermal comfort and energy. Building monitoring technology aim to improve the quality of life for occupants [7, 65, 68, 97] to maintain internal environment conditions [34, 98–102] and safety improvements [8, 103, 104].

BMS need to monitor real time building operation, record data, analyse information, and then address issues as soon as possible. Jang et al. [105] have described a methodology of real time monitoring. They considered that an application for a building monitoring system is divided into three parts: data acquisition, data collection and data retrieval. Authors in [106] described a decentralised system of software agents to monitor and control office buildings. The proposed system is used and resulted in energy saving and increasing customer satisfaction. In addition, LonWorks technology [107] have proposed a typical smart office building system containing an electrical network and a number of electrical monitoring and control devices.

Many companies are active in carrying out research to produce products for hardware and software monitoring of BMSs. A review of current products and companies is presented in Appendix A.

### 2.6.1 Monitoring Technologies

Development of hardware and software technology have improved monitoring efficiency of environments in and around buildings [105,108]. Hernandez et al. [68] presented the hardware and software design and implementation of a low-cost, wearable, and unobstructive intelligent accelerometer sensor for monitoring of human physical activities. Their proposed model is to improve the healthy lifestyle of people in buildings. The model is used to monitor daily activities and classify them. Tsai et al. [109] recommended that research needs to improve efficiency, accuracy, and power usage of monitoring devices. They have constructed practical designs to reduce the standby power consumption of PIR based power usage for lighting devices. Normal PIR lighting switches are active based on motion detected. The authors in [109] have designed a PIR able to reduce the consumed lighting power from 3 Watt to 0.004 Watt for each detection/activation.

Building activity monitoring is a challenging part of BMS. The tasks include appropriate sensor selection, collection and storage of information, and creation of application models to analyse and control [5]. Kasteren discussed sensors used in activity recognition. The sensors are reed sensor, float sensor, shake sensor, pressure mat, temperature sensor (ambient sensor), PIR, camera, wearable sensor and Radio Frequency Identification (RFID) [93].

### 2.6.2 Data Acquisition Technology

For the past few years, data acquisition via Wireless Sensor Network (WSN) has been applied in various fields such as physiological signals [110], healthcare [111] and intelligent transportation systems [112]. Chuo et al. [110] have proposed mechanically flexible wireless multi sensor for monitoring human physical activities. Tapia et al. [111] designed the interconnection of several networks from different wireless technologies, such as ZigBee or Blue-tooth for a tele-monitoring system for remote healthcare for dependent people at their homes.

A comprehensive survey of the architectures and technology of wireless home automation networks has been conducted [113]. In that survey ZigBee and other types of WSN such as internet protocol version 6 (IPv6) over Low power Wireless Personal Area Networks (6LoWPAN), Z-Wave, INSTEON and Wavenis are compared. The 6LoWPAN has the capability to solve problems that occur in network configuration and large address spaces. Z-Wave allows reliable transmission of short messages from the control unit. INSTEON is WSN that combines Radio Frequency and power line for the sender, which can also act as receiver or relayer in communication wireless. The other popular system is Wavenis, it is a protocol stack and operates at 4.8 kb/s and 100 kb/s [113]. Figure 2.5 shows that the latency results for data rate transmission are best with ZigBee. Nevertheless, 6LoWPAN has demonstrated good performance in data rate transmission [113].

In recent years, many studies about the capability and innovation scheme of ZigBee for improved connectivity in WSN [114] have proposed new approaches. ZigBee based on IEEE 802.15.4 protocol offers advantages such as low cost and low power [45, 115]. There are still opportunities to explore new approaches, design and concepts to optimise WSN for connectivity performance, unattended monitoring of a wide range of environments and real-time measurements [116].

Terada in [117] used a ZigBee sensor network for data acquisition and monitoring. The proposed monitoring system is configured using a ZigBee module to collect data from thermocouple sensors and then send data to a base station PC. Time event data is recorded and software is used to check and monitor real time data. The WSN architecture and topologies are configured as environmental monitoring contain ZigBee coordinator, router, end device, sink node, internet and user [3, 118].

The authors in [104] used the ZigBee embedded system to improve industrial safety quality. The proposed system provided wireless synchronized measurement and energy monitoring, temperature monitoring, $CO_2$ concentration, length filtering, ground vibration sensing, weight grading and electricity sensing. Ohtsuka et al. [119] proposed localization scheme on ZigBee WSN to carry out for networks and comprising sensor nodes. Yanfei et al. [120] improved the design using a ZigBee WSN on coordinator process.

Authors in [118] designed a multi-parameter monitoring system using ZigBee wireless communication technology. This system collected data for temperature and humidity monitoring. In another application [104], ZigBee is used in an industrial building to improve safety level of daily operation in order to achieve

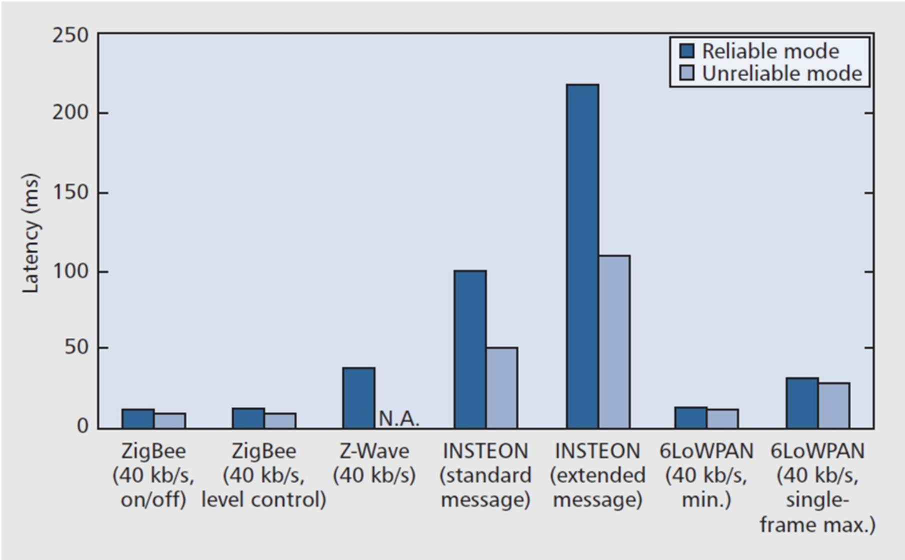



Figure 2.5: Expected latencies of transmission data rate between ZigBee, Z-Wave, 6LoWPAN and INSTEON [113].

industrial safety quality. As a next generation green home system, authors in [121] have demonstrated their smart home energy management systems based on ZigBee is able to control consumer homes wirelessly, manage home energy and adjust the environment automatically.

## 2.7 Computational Intelligence Techniques

Applying Computational Intelligence (CI) techniques in BMS is the subject of interest for many researchers. Considering the complexity and intuitive nature of the information collected from sensors, CI techniques would be able to process the information more effectively.

### 2.7.1 Clustering Techniques

Application of clustering techniques have contributed to intelligent building system to achieve energy efficiency, comfort satisfaction and improve lifestyle for occupants. A study in [122] has used clustering techniques on datasets for abnormal power consumption detection based on people's behaviour. Byun in [123] has proposed an Energy-efficiency Self-clustering Sensor Network (ESSN). The experimental results show that combining ESSN with a Node Type Indicator based on Routing (NTIR) protocol can produce a power saving of about 16-24%. Nazerfard et al. in [53] tried to improve mining algorithms using clustering techniques. They proposed Discovery of Temporal Features and Relation of Activities (DTFRA) using K-means clustering technique. The proposed approach is tested using four months of data collected in a smart home and predicted future human activities. Classification methods using fuzzy clustering algorithms have been proposed in [83]. A combination of an adjustable fuzzy clustering algorithm and probabilistic ANN has made recognition more efficient because it easily learns additional information based on features extracted from the wearable sensors data. Wang et al. [124] employed K-means clustering to improve apriori-kl algorithm to mine users' behaviour so that it could more effectively detect anomalous activities in a database. Rashidi et al. [33] in their system for discovering and tracking activities, used a clustering sequences to compare the similarity of sequence for recognising activities of people in CASAS smart environment.

### 2.7.2 Fuzzy Inference System

Application of Fuzzy Inference System (FIS) in BMS have shown improved performance in various approaches as listed below:

- Generating fuzzy rules to control power consumption or predict thermal behaviour and user's satisfaction [14, 42, 79, 98, 125],
- Developed the controller system using fuzzy controller-agent for HVAC [101]
- Fuzzy association for implement the cooperative analysis between dependent factors (individual-environment and energy) [78, 80], and
- Combined the fuzzy logic system with data mining technique for recognizing human activities [54, 72].

A study in [100] used FIS to improve BMS to tackle problems of non-linearities in controlling indoor thermal comfort. Fuzzy PMV /PPD are proposed using a Takagi-Sugeno (TS) fuzzy model to control HVAC to maintain the indoor comfort level. The model is tested and the results showed that it could achieve substantial carbon and energy savings and comfort satisfaction in a building. A study in [126] used a Mamdani-type FIS to monitor heat insulation of concrete in residential buildings.

Authors in [127] used a Genetic algorithm-based fuzzy-PID control methodology [128] to improve energy efficiency for BMS. Their system has demonstrated that it is able to save energy costs by about 51.2% during summer and 67.8% in winter. Authors in [129] used the same method as in [127]. Their system effectively saved energy costs of about 44% and 63% of per day average during summer and winter.

### 2.7.3 Artificial Neural Networks

In relation to BMS, Chengli Li et al. [130] proposed intelligent control of air-conditioning using the self-learning ability of Artificial Neural Networks (ANN). The model takes several factors of human thermal comfort as input and used a PMV value as a control target. Several other studies [99, 131, 132] also used PMV index as input data and used ANN to estimate the input data with varying degrees of efficiency.

An ANN is demonstrated capable of Predictive Building Energy Optimization [25], where a total of nine variables are used as the input parameters of a predictive model such as daily average dry-bulb temperature, daily average wet-bulb temperature, daily global solar radiation and daily average clearness index,solar aperture, daylight aperture, overhang and side-fins projections, and weekdays or weekend. In another study, Kazanasmaz et. al. [24] used a three-layer ANN model of feed-forward and included thirteen variables (date, hour, outdoor temperature, solar radiation, humidity, UV index and UV dose, distance to windows, number of windows, orientation of rooms, floor identification, room dimensions and point identification) as input for a power prediction model. The model performance is almost 98% accurate.

### 2.7.4 Adaptive Neuro-Fuzzy Inference System

Studies in [102, 133] used an Adaptive Neuro Fuzzy Inference System (ANFIS) to improve the overall performance of heating, energy efficiency and thermal comfort systems. ANFIS allows other method to be combined in order to develop building intelligence systems. For example, the least square method and the backpropagation algorithm are used with ANFIS to help a system to recognise the variables for estimating the average air temperature in multi-zone space heating systems [102, 133]. Experimental results have shown that the proposed system is able to improve the overall performance of heating systems, saving energy and improving building thermal comfort. Ari et al. [134] conducted studies about indoor comfort and energy optimisation. ANFIS is applied to control the environment to optimise the occupant thermal comfort. Using the ability of ANFIS in control system, Haqras et. al. [12] developed Intelligent Control Energy (ICE) for energy management of big commercial building. ICE has demonstrated significant energy cost reduction without reducing customer comfort levels in commercial buildings.

## 2.8 Discussions

Based on the review presented in this chapter, it is found that many studies have paid attention to achieving energy efficiency and building comfort satisfaction. There are a number of studies about these and related works on hospital, home, commercial building and office building. The literature shows that overall building energy usage is associated with building characteristics, occupant's behaviour and climate. Therefore, with these challenges, most studies related to building and comfort optimisation are focused on energy behaviour based on human activities in the building.

From the review presented in this chapter, it appears that a building equipped with a human activity recognition system can support a BMS to achieve energy efficiency and improve the occupants' comfort. As discussed in Section 2.7, applying CI techniques have proven significant contribution to increase performance of the BMS to achieve energy efficiency. As reviewed in 2.4.1 some of the studies have only been carried out on the user profile, but they have not considered linking the user behaviour profile to contribute to BMS in order to achieve energy efficiency. Therefore, the individual profile that contains information needs for support BMS to optimise energy consumption and occupants' comfort is important. Application of data mining techniques in human recognition system is necessary to identify different user profiles. More research is required to extract user profile based on user activity to achieve Building Energy Efficiency, so that the building environment can be automatically adjusted in line with user preferences.

# Chapter 3

# Experimental Architecture

## 3.1 Introduction

Office workers' performance and comfort in an office environment can depend on temperature, ventilation and lighting. Their power usage depends on the way they have adjusted their environment to optimise these characteristics, together with their usage of their PC. We must therefore consider the way in which one could monitor these characteristics (of both the worker and their environment) in order to log the data. The data capture and logging must be as unobtrusive as possible in order to avoid affecting the behaviour of the worker, and should not increase their stress.

Pervasive sensing technologies in intelligent environments offer unprecedented opportunities for providing intelligent office monitoring and self management to change office conditions that are directly suited to user satisfaction levels. As part of our investigation to understand an office worker's behaviours and ultimately optimise the energy usage, a monitoring system is required. The proposed experimental system architecture is explained in this chapter. Investigated techniques in data collection system are also presented in this chapter.

The remaining parts of this chapter are organised as follows; in Section 3.3 an overview of the proposed system architecture is presented. In Section 3.4, real data collection system is explained. In the Section 3.5 an overview of the development of a simulator for an office environment is described. The simulator is created to generate equivalent patterns to a human office worker. A discussion is provided in Section 3.6.

## 3.2 Cost of Hardware and Installation

Environmental control system development methods require proper investment and the use of the latest technologies to achieve increased energy consumption efficiency. The cost of installation and equipment and system operation is essential in order to achieve energy efficiency. Therefore, this study emphasizes upon using the latest equipment technology, which has the characteristics of low power consumption, durability, and easy maintenance.

The development of the data collection system in this study used cheap electronic and wireless technology components that have these characteristics. Developing intelligent office systems requires researchers to modify and design all incorporated electronic components according to WSN as shown in Table 3.1.

The constructed system was able to measure and collect the necessary information from an office environment. In order to enhance the ability of intelligent office systems, so that information can be recorded to a central database, a computer that costs approximately £450.00 and Matlab software are required. This

Table 3.1: The total estimated cost of hardware and installation for a single office monitoring system.

| Number | Component and Task | Cost Price |
|---|---|---|
| 1 | PICAXE AXE210 Connect Board (4 units) | £50.40 |
| 2 | IC MAX3232CPE/SP3232EUCP | £2.40 |
| 3 | PICAXE-18X microcontroller (4 units) | £14.40 |
| 4 | XBee Module (4 units) | £69.90 |
| 5 | Light Dependent Resistor | £1.00 |
| 6 | Humidity Sensor (HIH 4000001) | £24.00 |
| 7 | DS18B20 Temerature Sensor ( 2 units) | £4.80 |
| 8 | Pressure mat | £8.00 |
| 9 | Magnetic switch sensor (2 units) | £5.00 |
| 10 | Passive Infra-Red sensor | £4.00 |
| 11 | Miscellaneous and installation works | £20.00 |
| | TOTAL | £136.70 |

will improve the ability of the intelligent office system to process collected data for earlier predictions, decision-making processes, and human activity recognition learning, in order to achieve the purpose of the study. The system should also support software developed to ensure the accuracy of the information recorded at any time.

## 3.3 Proposed System Architecture

We may classify the needs of our experimental system for an office as the ability to detect occupancy, record the ambient conditions (and any worker responses to them), and to identify activities. Occupancy detection may be considered to be covered by PIR motion sensors and door entry sensors. PIR motion sensors are low cost and readily available, though their typical activity thresholds mean that a worker quietly reading or writing a document would not cause them to signal. Door entry sensors are relatively unambiguous. However, a visitor to an office would also trigger the open-shut sequence, so the office occupancy could not necessarily be deduced simply from the door. Thus occupancy needs to be deduced from a combination of these signals together with additional information provided by other sensors.

The ambient conditions can be recorded via temperature sensors (inside the office and outside the building) and light sensors. It is possible to identify when a person is actually typing at a PC, or moving a mouse, by adding a background software agent. If a worker is at a desk reading a report, the PC activity might halt for a considerable time, and the PIR might not trigger. An additional Boolean signal generated by a pressure sensor in the worker's seat could indicate when they are at their desk.

The proposed system architecture to measure office worker's activities and energy usage is shown in Figure 3.1. The proposed system comprises the following data collection components:

- sensors to measure user's activities and environmental properties,
- communication system to transfer collected data into a central database,
- PC monitoring application agent,
- a central database to store collected data.

The proposed system has also incorporated:

- behaviour identification and user profiling,
- computer usage, lighting and heating controller.

More details about the functionality of the above units will be provided in subsequent sections.

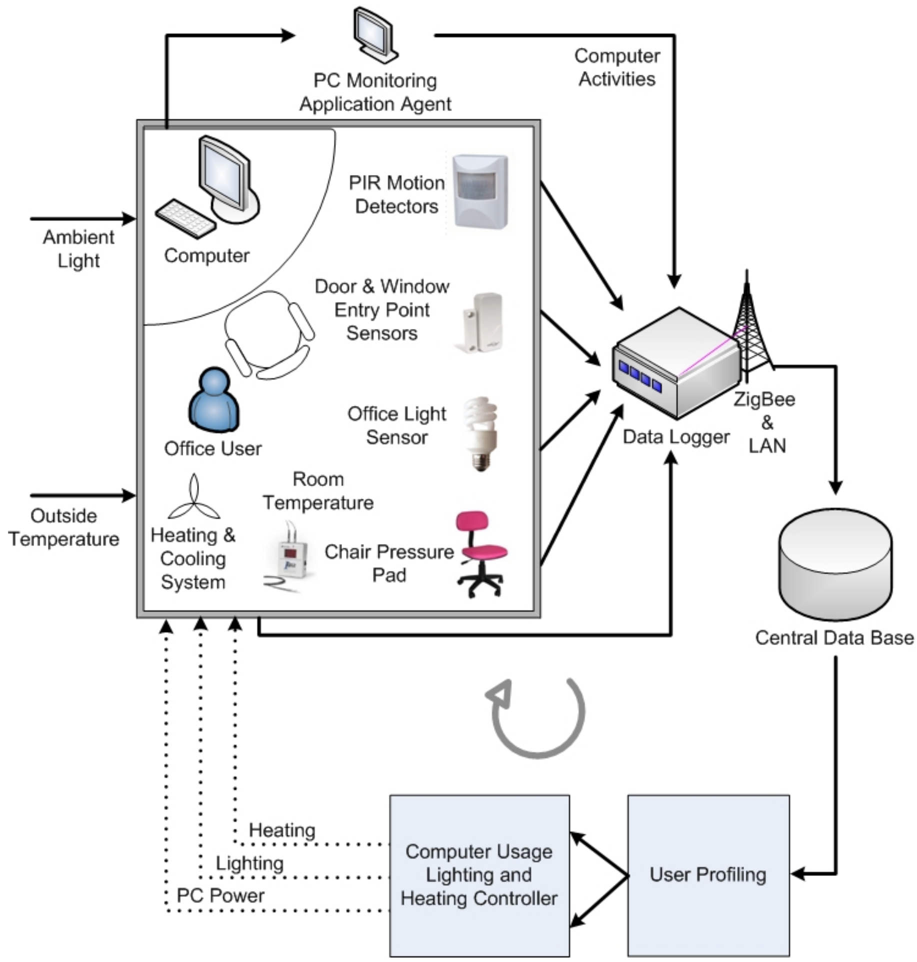



Figure 3.1: Proposed system architecture for intelligent office environment.

### 3.3.1 Sensors

The ambient conditions are recorded via temperature, humidity and light sensors (inside the office and external to the building). The responses of the office worker are also captured via his/her activities in the office environment. The most commonly used sensor types in smart environment are as follows [70]:

- PIR Sensor: PIR sensor is activated when detecting movement.
- Magnetic switch sensor: This sensor is suitable for door/windows entry. A magnetic field make the switch changes state from opened to closed and vice versa.
- Temperature sensor: There are a variety of sensors to measure the ambient temperature. For example, DB18B20 is a solid-state sensor commonly used to measure the environmental temperature. It operates in the range of temperature $-55°$ to $125°$ and can be powered from a data line with a range of $3.0\,V$ to $5.5\,V$.
- Humidity sensor: Humidity sensors, otherwise known as hygrometers, measure the relative humidity of the air. This is based on both air temperature and moisture. This dependency is a function of air temperature fluctuations. For example, the humidity sensor HIH-4000-001 is a solid-state sensing element with on-chip integrated signal conditioning. It is calibrated at $5\,V$ and $25°$.
- Light intensity sensor: The light intensity sensor is also know as a photoresistor which has the property of reducing resistance with increase in light intensity, allowing an increase in flow of photo-current.

- Electrical current sensor: A non-contact sensor is used to record the activity of electrically operated devices by measuring AC electrical current, and
- Pressure pad sensor: This type of sensor is mounted on chair to be able to pick the difference in the pressure on the chair.

A description of the sensory devices employed in the proposed monitoring system is given in Appendix B.

### 3.3.2 Communication System

To communicate the information gathered from sensors to the data logger system, both wired and wireless communication systems are investigated.

#### 3.3.2.1 Wired Sensor Network

In the first phase of our development, the data collection system is developed based on a wired system. The wired sensor network is created using Pico technology [135]. Two units of pico-log 1216 aided by one unit of pico-scope 2203 are used to create a Wired Sensor Network.

#### 3.3.2.2 ZigBee Wireless Network

In order to make the data collection system more flexible and less intrusive, a Wireless Sensor Network (WSN) is developed. The developed WSN is based on ZigBee wireless technology. The ZigBee network is proven to be a suitable technology for our application mainly because it is a low date rate, low power consumption and low cost wireless standard.

ZigBee is a wireless networking technology developed by the ZigBee Alliance for low data rate and short range application. According to IEEE 802.15.4 standard, the full function device can work as ZigBee coordinator or router in ZigBee network and operates in 868 MHz, 915 MHz and 2.4 GHz bands. These frequency signal bands are available in Europe, North America and worldwide respectively [113]. The application of ZigBee wireless network in many industrial control, electronic product and house automation are reported in [42] [119] and [136].

To develop a WSN to collect information from various sensors, the network is developed using XBee modules. XBee [137] is a wireless communication module that Digi (`http://www.digi.com`) built to the 802.15.4/ZigBee standard. Each node used PICAXE-18X micro-controller to program the node functionalities.

#### 3.3.2.3 XBee Testing and Configuration

XBee can be configured and tested using AT commands (to run programs and commands at scheduled times) via the X-CTU terminal interface program [137]. X-CTU is a Windows based application designed to interact with firmware. Figure 3.2 shows the process of reception of information from sensor nodes which function as the end device on WSN. According to Figure 3.2, the information received consists of room number, sensor ID and decimal value or states of sensor.

#### 3.3.2.4 Technical Considerations for the Project

The ZigBee wireless sensor network is designed based on both Star and Tree network topologies which includes sensor node, router and coordinator nodes. If one office is more than 20 meter away from the coordinator node, a router is used

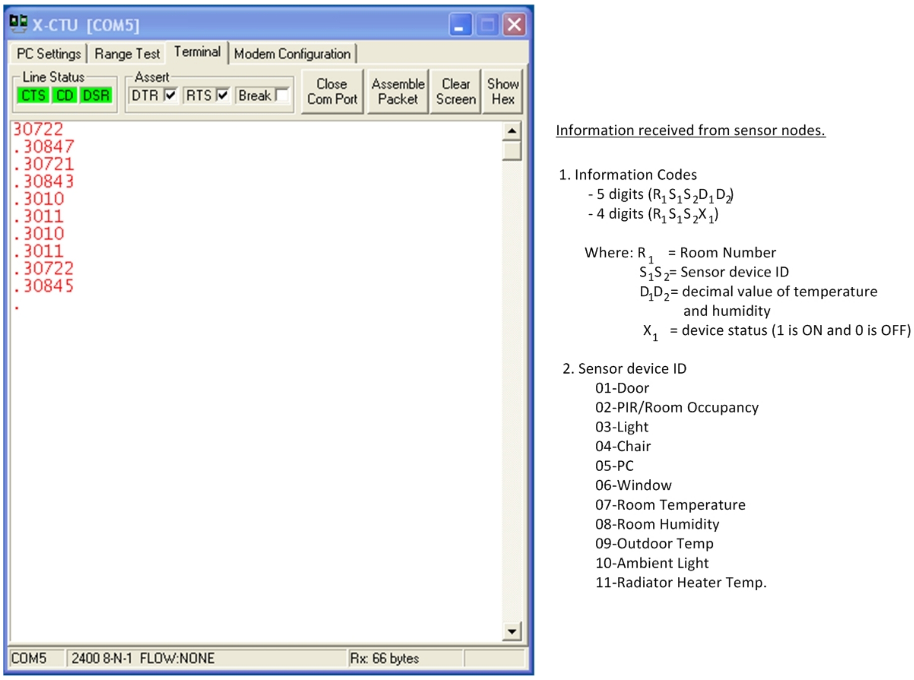



Figure 3.2: Screenshot of terminal tab of X-CTU application showing information received from sensor nodes.

to extent the physical range of a ZigBee network. [138]. The coordinator node is responsible for network address allocation, to collect data sent from sensor nodes and communicating with PC in base station through a serial port. Figure 3.3 illustrates the overall structure of the proposed wireless sensor network. A BASIC program in PICAXE Micro controllers manage XBee communication. More details about the sensor node devices and their interfaces are provided in Appendix C.

In order to reduce battery power consumption in sensor nodes, previous values of sensor data are used to compare with the current value of the sensor reading. This technique forces a XBee module to sleep mode and it only wakes up to send the new information if the current value is different from the prior value. Sensor nodes are programmed to produce information in a binary format, except those

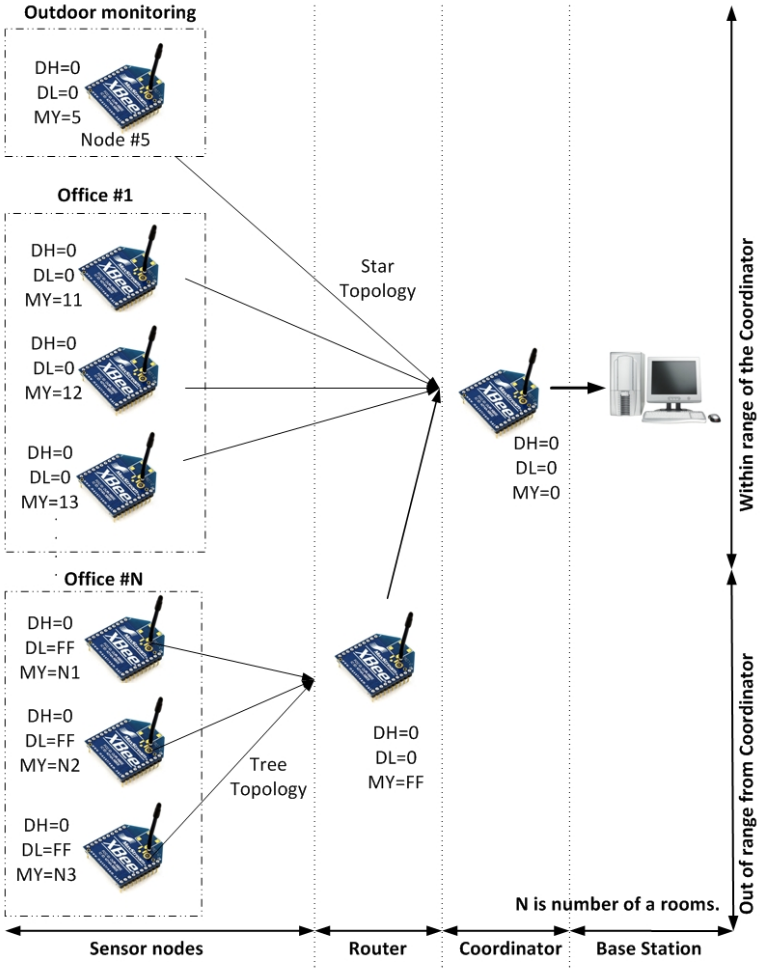



Figure 3.3: The structure of ZigBee wireless network.

for ambient light, temperature and humidity. In order to identify the data for each sensor and room location, a unique sensor ID within the data packet is used to transmit over the WSN.

Sensor data collected from PIR and chair pressure pad sensors are intermittent signals. For example, a person sitting on a chair adjusting their position could trigger the pressure sensor. This does not represent the occupancy or working pattern and it is better not to include the data for any further processing. To use the information gathered from PIR and pressure pad sensors, the PICAXE

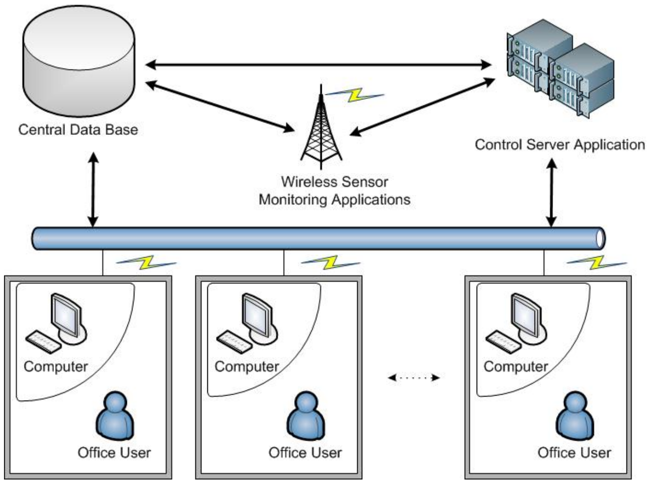



Figure 3.4: Proposed PC monitoring using internet technology.

micro-controller is programmed to add a delay in the transmission of changes. 10 minutes delay for PIR and 5 minutes for pressure pad sensors are applied. A sensor node #0 as network coordinator consists a XBee module and a SP3232E IC is connected to the server (base station). The software on the server is responsible for recording data, monitoring the user activities, analysing and predicting.

### 3.3.3 Personal Computer Monitoring Application

Information regarding the office PC is collected using a monitoring application agent installed on the PC to be monitored. The architecture of PC monitoring using internet technology is proposed as shown in Figure 3.4. A monitoring software agent logs keyboard and mouse activities. These activities are recorded in a database along with the rest of the data collected from the WSN.

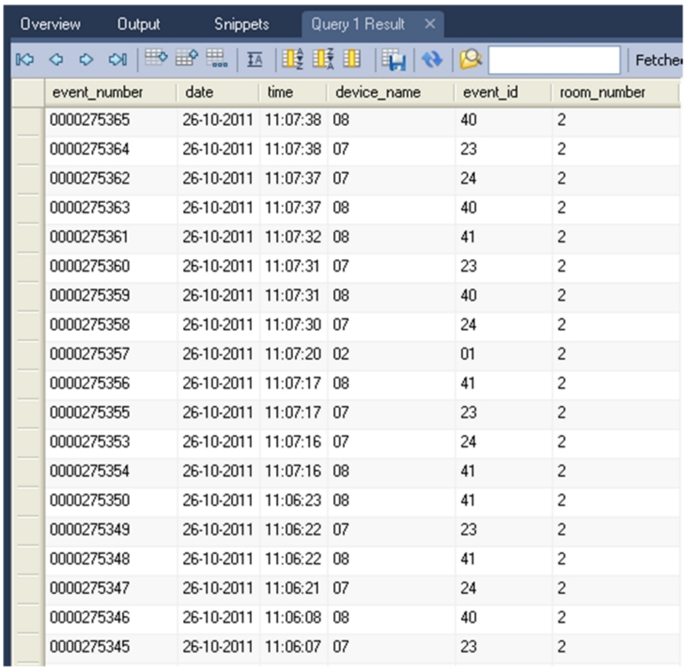
The important part of the monitoring application has to detect when mouse or keyboard events happen. After being able to detect keyboard and mouse events to indicate they are in use, it needs to be determined when the keyboard and mouse are no longer in use. The best way to do this would be to have a time period where if no mouse or keyboard events have been detected then they are deemed to be no longer in use. Activating a timer after detecting the mouse or keyboard is in use would make this possible. It would have to be possible to reset the timer though as if any mouse or keyboard events are detected while they stated to be in use then the time range to deeming not in use needs to be restarted. This would mean that if the main application is doing a task then the timer would be stopped until that task is completed and vice versa.

The PC monitoring application needs to communicate with the control server application in both directions. It needs to send a message to the control server application when the PC has gone idle or change state, informing the control server application that the PC's state has changed.

### 3.3.4 Database

A database records events triggered by the PC software agents and WSN. Figure 3.5 shows a sample of the database data format. The database logs individual events for any further analysis. In the database design, the database contains tables for device usage events, PC MAC address, PC statuses and sensor statuses. The following tables are used for managing storing data in the database system:

1. **Device Usage Events** - This table is used for storing changes in the states of the wireless sensors and PC. The primary key is the event number which auto increments from zero. The reason the date and time is not used as the primary key is because multiple events could happen at the same moment so would be recorded with the same date and time values. This breaks




| event_number | date | time | device_name | event_id | room_number |
|---|---|---|---|---|---|
| 0000275365 | 26-10-2011 | 11:07:38 | 08 | 40 | 2 |
| 0000275364 | 26-10-2011 | 11:07:38 | 07 | 23 | 2 |
| 0000275362 | 26-10-2011 | 11:07:37 | 07 | 24 | 2 |
| 0000275363 | 26-10-2011 | 11:07:37 | 08 | 40 | 2 |
| 0000275361 | 26-10-2011 | 11:07:32 | 08 | 41 | 2 |
| 0000275360 | 26-10-2011 | 11:07:31 | 07 | 23 | 2 |
| 0000275359 | 26-10-2011 | 11:07:31 | 08 | 40 | 2 |
| 0000275358 | 26-10-2011 | 11:07:30 | 07 | 24 | 2 |
| 0000275357 | 26-10-2011 | 11:07:20 | 02 | 01 | 2 |
| 0000275356 | 26-10-2011 | 11:07:17 | 08 | 41 | 2 |
| 0000275355 | 26-10-2011 | 11:07:17 | 07 | 23 | 2 |
| 0000275353 | 26-10-2011 | 11:07:16 | 07 | 24 | 2 |
| 0000275354 | 26-10-2011 | 11:07:16 | 08 | 41 | 2 |
| 0000275350 | 26-10-2011 | 11:06:23 | 08 | 41 | 2 |
| 0000275349 | 26-10-2011 | 11:06:22 | 07 | 23 | 2 |
| 0000275348 | 26-10-2011 | 11:06:22 | 08 | 41 | 2 |
| 0000275347 | 26-10-2011 | 11:06:21 | 07 | 24 | 2 |
| 0000275346 | 26-10-2011 | 11:06:08 | 08 | 40 | 2 |
| 0000275345 | 26-10-2011 | 11:06:07 | 07 | 23 | 2 |

Figure 3.5: Sample of database data format.

the rule of a primary key being a record identifier. The event number gets around this issue and by using a big INT variable as its data type it can fit billions of entries so there is no worry about it being filled. The rest of the fields in this table enable it to record office status and ambient conditions.

2. **PC MAC Address** - PCs on a network must have a unique identifier which makes it appropriate as a primary key. This ties a MAC address to a PC name. By simply looking up the name of the PC, the control server application can find its MAC address on the local area network.

3. **PC Statuses** - This table stores the current status of any PCs that have registered with the control server application. This table is used by the control server application to monitor PC activities. When a message is sent to the control server application by the wireless sensor monitoring

application or the PC monitoring application indicating that a change of state in the PC or its environment has occurred, the name of the PC that sent the message or the PC the wireless sensor is monitoring is recorded. The control server application can then use that PC name to retrieve its current state from this table. The idle status indicates whether the PC is currently in use, the room number indicates in which room the PC is located and the off status indicates if the PC is currently in its power management mode, turned off or the monitoring application is not running.

4. **Sensor Status** - This table stores the current status of all the sensors monitoring an office in the proposed system. The sensor ID identifies each sensor and acts as the primary key. The sensor ID is made up of two parts where the first part represents type of sensor. For example, a pressure pad sensor in room number 2 will have the sensor ID 2-04. The sensor status field can indicate two things depending on the type of sensor. If a sensor determines something is active or not, such as a door being open or closed, it will have a value of zero or one. If the sensor is one that measures something, like humidity or temperature, then the value logged is the numerical value that the sensor has recorded. Whenever a change in a sensor state occurs, the wireless sensor monitoring application will update this table using the ID of the sensor.

Figure 3.6 shows Enhanced Entity Relationship diagram of database. The *Device Usage Events* table has a many to one relationship to *Sensor Statuses* table and *PC Status* table. The *PC Status* table has a many to one relationship to *PC MAC Address* table. *Sensor Status* table has one to many relationship

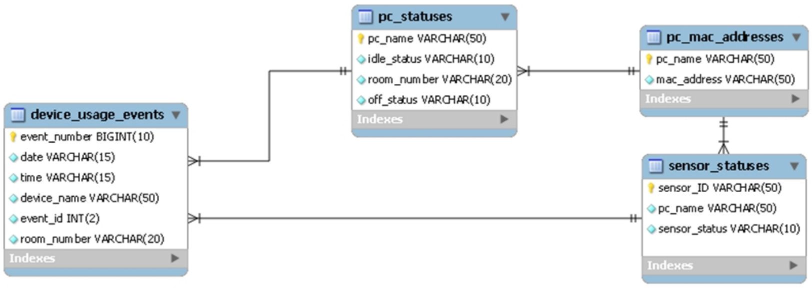



Figure 3.6: A enhanced entity relationship diagram of database.

with *Device Usage Events*, which also have many to one relationship to *PC MAC Address*.

To check the status of all sensors and PC activities, an on-line web monitoring portal is also developed. The developed system called "iOffice" is accessible by all users on the network. The iOffice web portal is used to validate the data collection system and check that the sensory devices are working properly during monitoring. Figure 3.7 shows a snapshot of the web interface where a query menu is used to retrieve data by date and room number from the database. Retrieved date can be displayed either in graphs or database data format.

## 3.4 Data Collection

The office environment used as the testbed for our experiment are four academic staff offices at Computing and Informatics Building, Nottingham Trent University. Figure 3.8 shows the layout of the offices where the experiments are conducted. Collected data from offices include room status, power use of office equipment and ambient information. In total, each office is equipped with 10 sensors measuring different activities and properties of the testbed environment.

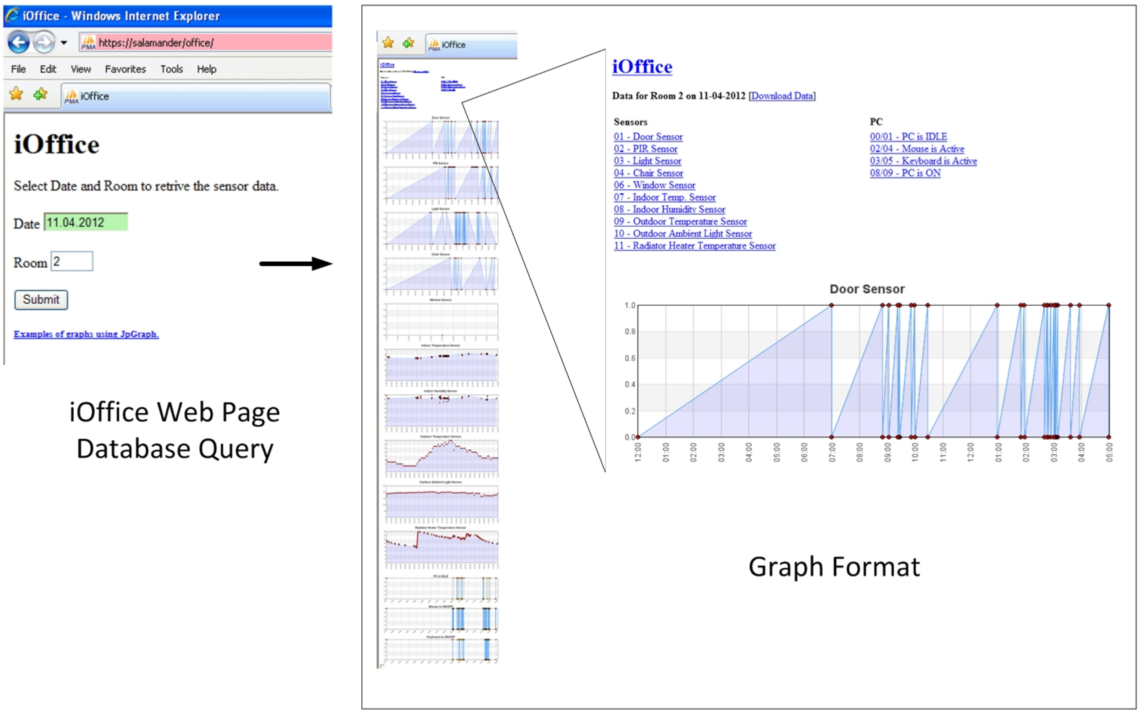



Figure 3.7: iOffice web interface.

The sensors include:

- Motion sensor (On/Off)
- Door entry point sensor (On/Off)
- Windows entry point sensor (On/Off)
- Pressure sensor measuring chair occupancy (On/Off)
- Room temperature
- Room humidity
- Outside temperature
- Room light intensity

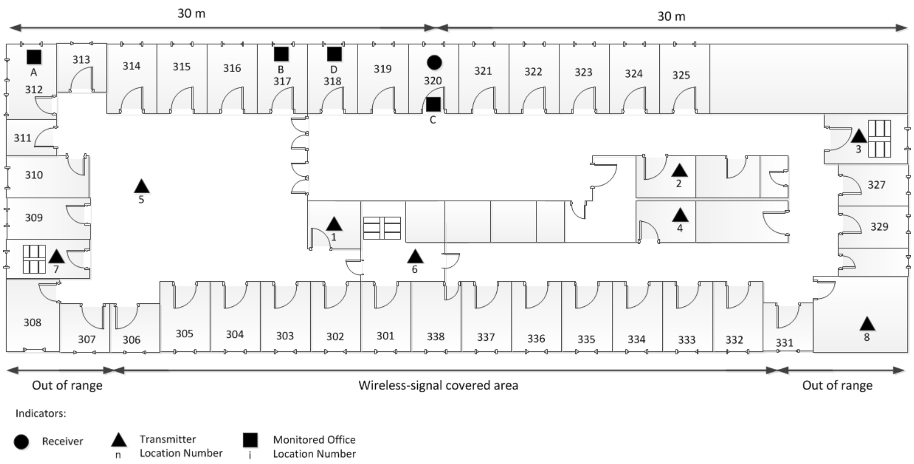



Figure 3.8: Testing the Zigbee RF properties in an office building environment.

- Ambient light intensity
- Radiator Heater Temperature

These sensors provide information regarding the occupancy and ambient conditions.

Table 3.2 shows summary of collected data during the course of the research. There are four office represented as A, B, C and D. There are four different users represented as user #1, #2, #3 and #4. We have seven data sets which is used for further analysis. In the rest of this thesis, we refer to these data sets as D1, D2, ... D7.

Figure 3.9 shows a snapshot of office signals representing ADW for user #2. Office door activities, room occupancy status, light activities, chair occupancy status, computer activities, window open/close status, radiator temperature, room temperature, room humidity, outdoor temperature and ambient lights are recorded. The ADW information for the same user over a longer period of time is shown in Figure 3.10. Data are collected over two phases of data collection. In

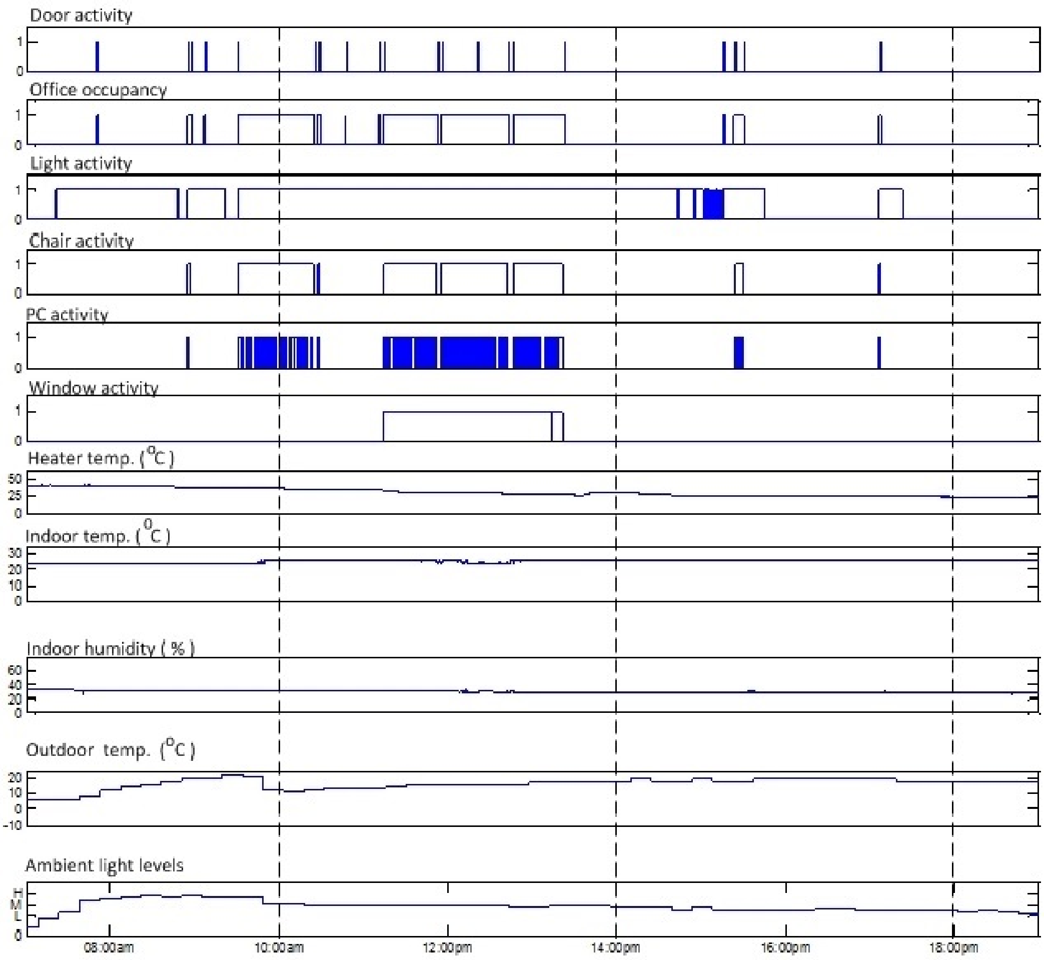



Figure 3.9: Sample of office signals representing user #2 daily activity in an office environment on 27-March-2012.

Table 3.2: Summary of Collected Data

| Data | User | Office | Start | End |
|---|---|---|---|---|
| D1 | #1 | A | 27-11-2011 | 17-12-2011 |
| D2 | #2 | B | 27-11-2011 | 17-12-2011 |
| D3 | #3 | C | 27-11-2011 | 17-12-2011 |
| D4 | #1 | A | 09-01-2012 | 07-07-2012 |
| D5 | #2 | B | 09-01-2012 | 07-07-2012 |
| D6 | #4 | D | 09-01-2012 | 01-04-2012 |
| D7 | #3 | C | 01-04-2012 | 07-07-2012 |

the first phase, datasets labelled as D1, D2 and D3 are gathered. In the second phase, D4, D5, D6 and D7 are collected.

## 3.5 Office Environment Simulator

Based on the real data collected from the office environment, a flexible simulator is developed which accepts different office worker profiles including expected office occupancy and computer usage. The simulator generates sensory signals which represents different activities and occupancy of the office environment. The validity of the simulator is verified by tuning the simulator parameters to occupancy data collected by sensory systems from real offices.

Developing the simulator will allow us to generate large data sets in a shorter period of time. This would also allow us to change the parameters and see their effects. By creating relatively extreme behaviour (which normally does not happen in real environment), we would be be able to test the modelling and prediction capability of the proposed system.

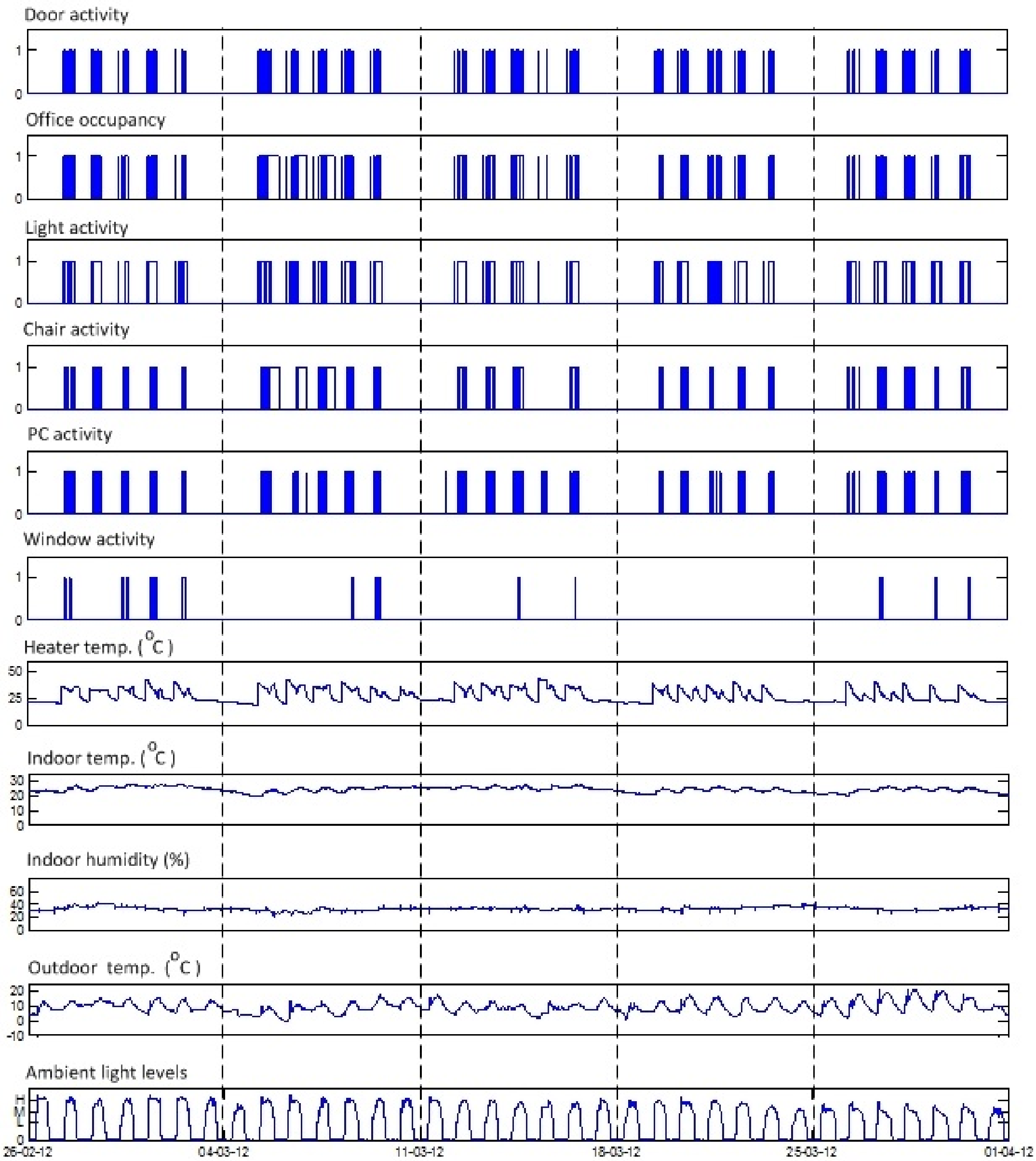



Figure 3.10: Sample of office signals representing signals recorded by the data collection system from 26-Feb-2012 to 01-April-2012.

### 3.5.1 The Development of Office Environment Simulator

The simulator starts from the assumption of an office occupied by a single worker, who carries out a range of activities during the standard working day. It also starts by assuming that information is being generated by (virtual) sensors reporting on aspects of the behaviour. This is an important aspect of the realism, in that it allows the simulator to generate data that needs the same sort of interpretation as the real office. If the simulator simply produces records "Reading", "Typing at PC", "Out of Office" for example, it would have little value in helping to produce algorithms pertinent to the real environment, where activation, say, of the chair sensor could be used as an indicator of "Reading" only if it did not overlap with the data indicating that the mouse is being moved.

This research uses stateflow to define the states for modelling each mode of operation. Stateflow is a powerful graphic method for supervisory logic problem and allow us simulate complex reactive systems based on finite state machines [139]. According to the stateflow process, developing an office environment simulator can be divided into five steps. Define the interface to Simulink, define and structuring the states, identifying state action and variable, specifying transition between states and generating the parameters to drive the activity. Finite-state machines (FSM) represent operating modes as states. For example, a computer activity can have states such as On and Off, while temperature level can have states such as High, Medium, and Low. To construct finite-state machines, Stateflow provides a graphical representation of a FSM where states and transitions form the system. Figure 3.11 shows an example which models as a FSM the states required to control the sensor activity of sensor simulator system.

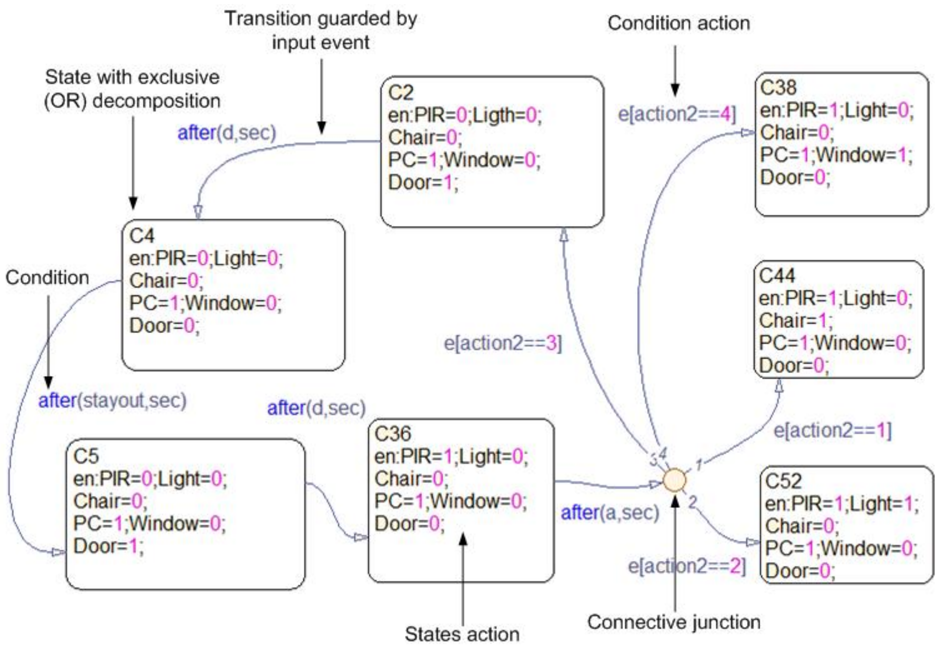



Figure 3.11: Stateflow chart.

For this reason, each of the macro-states in the simulator has to be characterised by its name, allowed transitions, duration, and whether it overlaps with others. The micro-states of the individual sensors are much simpler, and are as shown in Figure 3.12. Combining the individual sensor models to take into account the possible overlapping of events and parallel activations of some of the sensors leads to the Finite State Machine of the office shown in Figure 3.13.

In order to drive the behaviour of the simulator, it is necessary to create a model of human activities, with a simple set of actions characterised by a few parameters. Logic 1 or 0 are used as state values to represent sensor active or inactive in order to extract the user's behavioural signal in an office environment. For example, the simulation output signals of an office simulator which contains six different sensors signal are depicted in Figure 3.14. This is to allow us to model

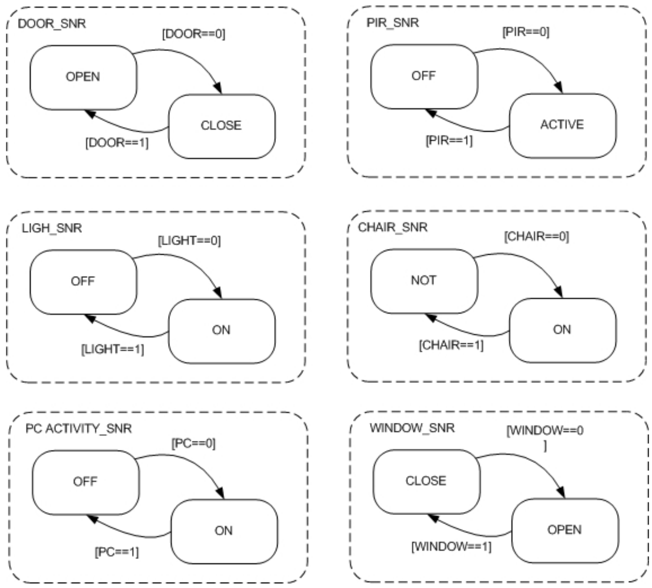



Figure 3.12: Individual state models for sensors.

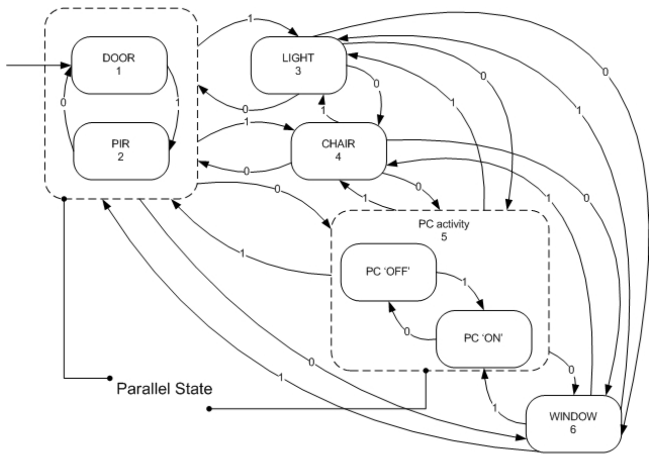



Figure 3.13: Finite state machine for the office simulator.

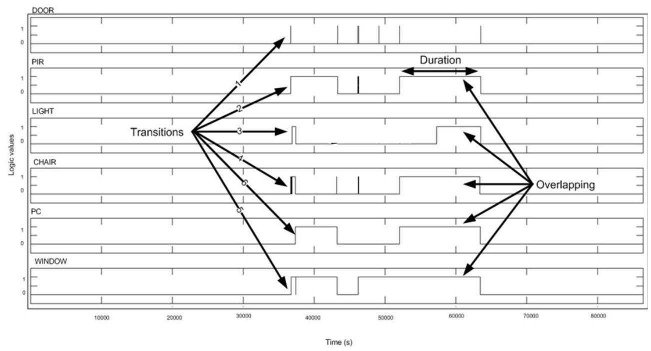



Figure 3.14: Sample of activity signals.

a range of different workers' routines with differing degrees of chaos. The human model generates a sequence of 'actions' lasting for differing amounts of times. In order to create the behaviour of a person doing an activity that is as natural as possible, the values of the parameters are generated using random numbers based on a normal distribution. The actions, their start times, durations and transitions from one to another then feed into the office model, so that the individual sensor state machines are enacted - some in parallel and some in sequence, depending on the action being considered.

### 3.5.2 Structure of Simulator System

The office environment simulator can thus be considered to be made of four parts: user parametrisation, office simulation, sensor simulation and output. The sequential structure allows the system run in AND and OR mode (see Figure 3.12). All states of the sensors use logic 1 and 0 to represent the active and inactive states. Figure 3.15 shows the block diagram of the office environment simulator. The output block of this simulator receives a trigger signal from the sensor simulation block, and all activity sensors depending on user activity are

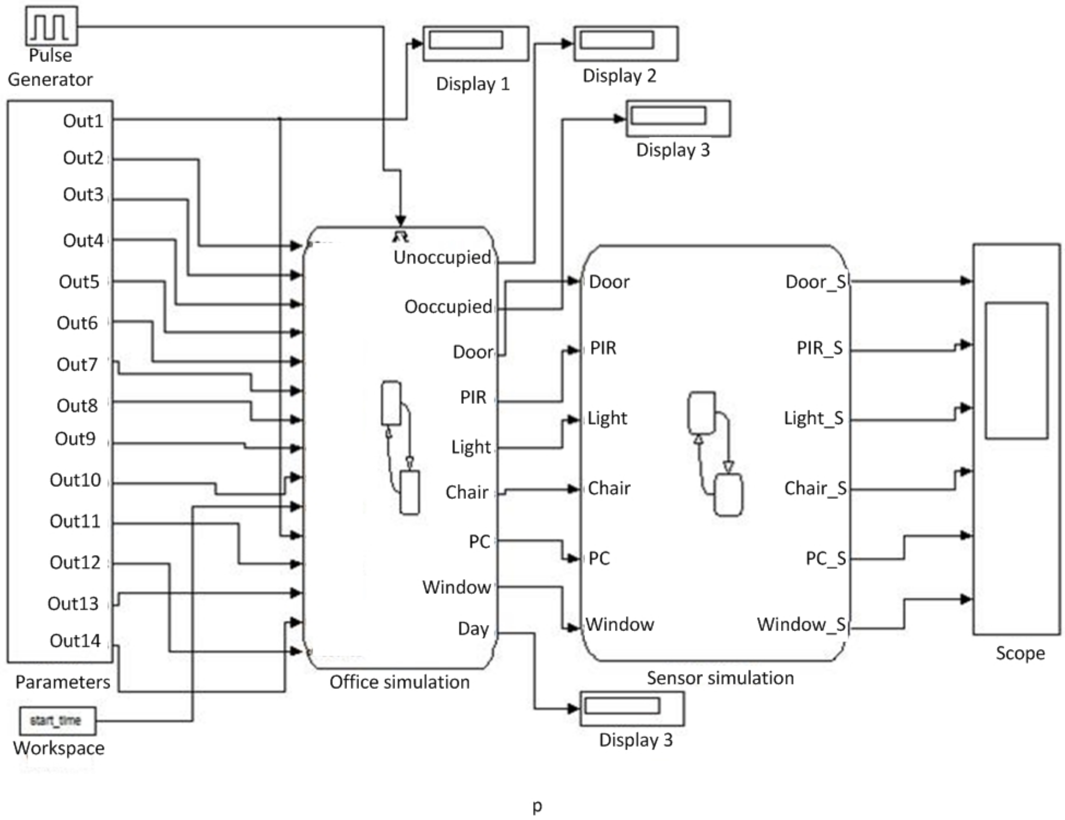



Figure 3.15: Block diagram of the office environment simulator.

processed into the office simulation block. The office simulation block is controlled by the user parametrisation block.

### 3.5.3 Parametrisation Process

Block parameters are defined as the input system of an office simulator, used to generate parameter values to control the simulation flow of following office users behaviour. Time controllers and action randomness decision are the main parameters that can be defined as person characteristics. The simulation system needs input values, such as start action, work time, leave work time, duration time of actions, and counters. In order to create a model of a person doing office

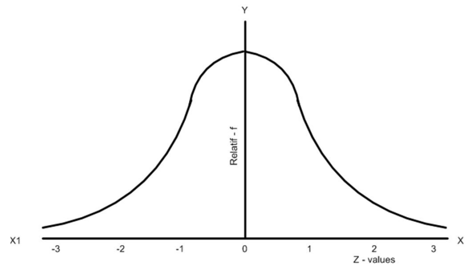



Figure 3.16: The standardised normal curve

activities, all parameter values are generated using random numbers based on a normal distribution. By generating parameter values using random numbers and uniformly random integers in the range [0, M-1], where M can be scalar or vector, it becomes possible to model the selection for transition, time activity and duration.

$$x_{k+1} = ax_k + c \bmod M \tag{3.1}$$

In Lehmer's random generators [140], a sequence integer can be defined as Equation 3.1, where it involves three integer parameters, $a$, $c$, and $M$ and an initial value, $x_o$, called the seed. Let's say, if a = 13, c = 0, m = 31, and $x_o = 1$, the sequence begins with 1, 13, 14, 27, 10, 6, 16, 22, 7, 29, 5, 3, and so on.

The graph shown in Figure 3.16 states that the standardized normal curve is obtained from the bell-shaped curve using the substitution,

$$z = \frac{x - \bar{x}}{\sigma} \tag{3.2}$$

where the $z$-value is a multiple of the standard deviation. The substitution value of z, converts the original distribution into one with zero mean and unit standard deviation. The equation of the standardized normal curve is:

$$y = \phi(z) = \frac{1}{\sqrt{2\pi}} e^{-}(0.5)z^2 \tag{3.3}$$

Generating parameter values, using a random number based on a normal distribution, makes all output random variables independent and identically distributed. If the M number is a vector, then its length must be equal to the length of the initial seed. In this case, each output has its own output range.

## 3.6 Discussion

In this chapter, an experimental architecture for the proposed office data collection was introduced. The architecture of an office simulator, used to generate more data for further analysis, was presented. The data is collected for distinctive users in different office environments. Experiments were performed to track the occurrence of regular user activities, in order to detect changes in an individual's pattern. The aim of these experiments is to establish monitoring sensors, communications, and data collection approaches in real office environments.

The data collection and performance of the Intelligent Office system could successfully monitor the activities of users, illustrate users activities in an activity's graph, in a start-time-duration format, link data across records and record data over long periods of time. Theoretical analysis, hardware and software design, and experimental environment testing showed that the ZigBee wireless sensor network system works effectively as a monitoring system.

In the remaining part of this thesis, data mining techniques will be used to extract user's preferences from the user's behavioural dataset, in order to construct different user profiles. The environment control system may be customised, so that office conditions may be regulated automatically, in line with individuals user profiles. This will allow energy efficiency and office worker comfort optimisation.

# Chapter 4

# Data Analysis Techniques

## 4.1 Introduction

The sequence of activities associated with a user in the workplace represents the pattern of their work. To identify similarities or dissimilarities between different days for a single user and compare that with the pattern of anther user, it is important to investigate different measures which could accurately represent the relationship between days and users. In the next section, a brief description of similarity measures used in our investigation are presented.

Measuring the dissimilarity (distance) between two binary sequence of equal length is the subject of interest in many research areas. A comprehensive survey of binary distance and similarity measures are presented in [141]. The authors compared 76 different dissimilarity and similarity measures and classified them through hierarchical clustering. The properties of binary vector dissimilarity measures are discussed in [142]. Distance and similarity measures for binary vectors/features are used in various fields including pattern matching and pattern recognition [143].

The remaining sections of this chapter are organised as follows: in Section 4.3 an overview of the statistical techniques are presented followed by an explanation of similarity/dissimilarity measures in Section 4.4. Discussion are provided in Section 4.5.

## 4.2 Ethical Issues

In the employment sector, there are ethical implications of employee and employer monitoring. Reviews by Mujtaba discussed current practices related to the ethics of monitoring [144]. They found that monitoring can be done both at work and at home via the internet. For example, issues such as privacy of work carried out and security. However, this study suggests that guidelines and policies should be created to identify the type of information collected, so that security and misuse of information can be controlled.

Referring to the study by Johnathan et al. [145], related legal and ethical issues of employees for improved organizational performance monitoring - this made the study's implementation of appropriate monitoring work as follows:

- Form data not related to personnel information and the importance of institutions. The information collected and recorded can not be used for any other purposes that are prejudicial to the people involved.
- Environment variables are involved only in relation to the goal of the study in order to improve and build an environmental control system, to save energy and increase the comfort of working in an office environment.

- Getting permission from the individuals involved - before the study is conducted.
- Using non-intrusive monitoring system.
- The monitoring system used does not reduce the level of security and affect productivity or comfort.
- Help to improve quality, administration and management.
- The monitoring equipments does not include camera.

## 4.3 Statistical Techniques

Statistical analysis based on duration of an activity is widely used in pattern discovery and human activity recognition [33,60,62]. Specifically, statistical based analysis is used in [53] to face the challenges of human activities recognition namely recognizing concurrent activities, interleaved activities and ambiguity of interpretation.

Consider a sample sequence of activities (e.g. using the chair) shown in Figure 4.1. The duration of an activity, $d$, is measured by the length of time an activity stays in a state. The duration variable, $d$, is only suitable to represent activities which have only two states $ON$ and $OFF$. The duration range of an activity can be in seconds, minutes, hours or even days. One activity can be represented as a sequence of $n$ durations $D = [d_1, d_2, d_3, ..., d_n]$. Standard statistical techniques can be applied to analyse the user behaviour patterns based on the duration of activities.

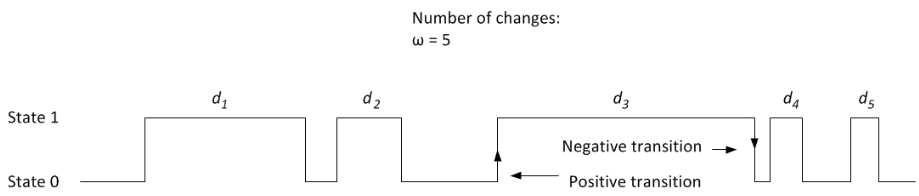



Figure 4.1: A sample sequence of an activity containing two states.

To perform the statistical analysis, the data collected from user's activities in an office environment is examined using several statistical techniques. Sum of duration, number of changes, consistency measures and approximate entropy measures are used to quantify the amount of regularity and the unpredictability of changes over time in ADW of a user in an office environment.

### 4.3.1 Consistency Measure

The consistency of user activity can be measured using Cronbach's $\alpha$. It is a statistical measure of internal consistency and it is mostly used to determine the reliability of a set of questions that are asked to respondents [146]. The level of consistency for an activity is according to the range of Likert scale. In this case, ranges of 0 to 1 are used to represent the level of consistency. Where, the lower level of consistency is if the value of Cronbach's $\alpha$ close to 0, while the higher level of consistency is if the value of Cronbach's $\alpha$ close to 1 [146].

To measure the consistency of the activities over different days/weeks/months, it is important that we present the activities in a comparable format. To present the data with a fixed dimension, total daily duration of an activity is used. Therefore, activities over a week can be presented as an $m \times n$ matrix, where $m$ is the number of weeks and $n$ is the number of days.

The Cronbach's $\alpha$ consistency measure can be calculated as:

$$\alpha = \frac{D}{D-1}\left(1 - \frac{1}{x_T^2}\sum_{n=1}^{D}\right) x_n^2 \tag{4.1}$$

where $D$ is the number of days, $x_n^2$ is the variance of duration in each day, and $x_i^2$ is the variance associated with the sum of entries value for each day ($D$), and $T$ is total number of days. $m$ is a number of rows and $n$ for columns in a matrix.

To evaluate internal consistency for the matrix, the Cronbach's $\alpha$ coefficient can be formulated as follows [146]:

$$\alpha = \frac{D\bar{d}/\bar{\mu}}{1+(D-1)\frac{\bar{d}}{\bar{\mu}}} \tag{4.2}$$

where $\bar{d}$ is average of $D(D-1)$, $\bar{\mu}$ is the average of all $D$ item variances $x_n^2$. According to [146], in the calculation of sum of variances of $x_n^2$ and $x_T^2$, there will be a problem due to different units. Therefore, a standardised coefficient, $\alpha_s$ is suggested, which is formulated as follows:

$$\alpha_s = \frac{D\bar{\beta}}{1+(D-1)\bar{\beta}} \tag{4.3}$$

where $\bar{\beta}$ is the average for $\frac{D(D-1))}{2}$ pairwise correlation coefficients between the summation values of duration activity of $D$.

### 4.3.2 Approximate Entropy Measure

This measure was initially proposed by Pincus [147]. Approximate Entropy (ApEn) is a measure that is widely used to provide an index representing complexity of data [148]. Chon et al. in [148] used ApEn to measure complexity of various signals. They argured that ApEn is able to find out the conditional probability of similarity for data in different sets of segment. Sattler et al. [149] used ApEn to analyse simulated signal data from noise in an electrical power system. The ApEn shows effectiveness to detect certain aspects of a system's behaviour.

Due to dynamic human behaviour influencing pattern formation, because of the effectiveness of the measure to measure the chaotic level of user activities, it is suggested to use ApEn in our research. We define a set of changes of an activity as a signal $D = [d_1, d_2, d_3, ..., d_n]$, where $n$ is the total number of data points for a period of time. According to study [150], ApEn required $m$ parameter as embedding dimension and $r$ parameter as a tolerance value, and then ApEn can be summarised as below:

1. Form m-vector, $D(1)$ to $D(n-m+1)$ can be defined by

$$D(n) = [d(i), d(i+1), ..., d(n+m-1)], n = 1, N-m+1. \tag{4.4}$$

2. The distance $Dis[d(n), d(j)]$ between vector $d(n)$ and $d(j)$ can be defined as below:

$$Dis\,[d(n), d(j)] = \max_{k=0,m-1}[|d(n+k) - d(j+k)|] \tag{4.5}$$

3. Given $n$, for $n = 1,\ N-m+1$, $C_r^m$ can be calculated as below:

$$C_r^m = \frac{V^m(n)}{N-m+1}, \tag{4.6}$$

where,

$$V^m(n) = Dis[d(i), d(j)] \leqslant r \tag{4.7}$$

4. According to Equation 4.6 and 4.7 respectively, a value of $\Phi^m(r)$ can be calculated by taking the natural logarithm for each $C_r^m$ and average it over $n$. Therefore, formula to find $\Phi^m(r)$ as below:

$$\Phi^m(r) = \frac{1}{N-m+1} \sum_{n=1}^{N-m+1} \ell n\left(C_r^m(n)\right) \tag{4.8}$$

The calculation of $\Phi^m(r)$ will complete with increase the dimension to $m+1$ and repeat steps 1 to 4.

5. Finally, the ApEn can be calculated for values for a finite data length of $N$ as formulated below;

$$ApEn(m, r, N) = \Phi^m(r) - \Phi^{m+1}(r) \tag{4.9}$$

where, $C_r^m$ is correlation dimension, and parameter $\Phi^m(r)$ is used to indicate the Eckmann-Ruelle entropy formula [147].

## 4.4 Distance Measures

Distance measures are used to give a natural notion of the quantitative distance between two objects or events. There are many types of distance measure defined in different subject areas. Distance measures denote dissimilarity, and this dissimilarity represents the discrepancy between the two objects or events. Therefore distance measures are opposite to similarity measures. Similarities are quite difficult to measure. To measure the similarities between two events or features, one can measure the distance which represent the dissimilarities between the events or features.

### 4.4.1 Linear Distance Measures

A binary sequence $X$ is defined as:

$$X = (x_1, x_2, \ldots, x_N) \tag{4.10}$$

where, $x_k \in \{0, 1\}$, $\forall k \in 1, 2, \ldots, N$. Let $\Re$ be the set of all N-dimensional binary sequences. The complement of binary sequence $X \in \Re$ is defined as $\bar{X} = I - X$ where $I \in \Re$ is a unity sequence with every bit sample equal to '1'. Given two sequences $A \in \Re$ and $B \in \Re$, the dissimilarity $D(A, B)$ or similarity $S(A, B)$ are defined. Let $M_{ij}(i, j \in \{0, 1\})$ be the number of occurrences of matches with $i$ in $A$ and $j$ in $B$ at the same time positions. Therefore, it can be easily deduced that $M_{00} = \bar{A}.\bar{B}$, $M_{11} = A.B$, $M_{01} = \bar{A}.B$ and $M_{10} = A.\bar{B}$. Using $M_{00}$, $M_{01}$, $M_{10}$ and $M_{11}$ many dissimilarity/similarity measures are defined. In Table 4.1 a sample of some of the proposed distance measures and their associated similarity measures for binary sequences are presented. The measures presented in Table 4.1 are not normalised.

Conventional definitions of distance measures include correlation, inner-product and Hamming distance based [151,152]. A study in [143] concluded that the conventional similarity coefficients depend on their type. According to this study, the main division is between the inner product based similarity measure and other measures which consider positive and negative matches. The Hamming distance measure is a simple measure of differences where the number of differences between bit values of corresponding bit positions in each sequence is calculated. For example for binary sequence $A = 1110000111$ and $B = 1111000011$ the distance will be 2. The Hamming distance measure would not take into account any dependency on the neighbourhood of the bits. In all measures mentioned above, local correlations are discarded. However for temporal binary sequences, the dependency and local correlation should be considered. A linear-time algorithm for Hamming distance is introduced in [153].

As noted in [152], when two binary sequences are compared, some degrees of fuzziness in measurement of distance are desirable. Consider portions of three binary sequences representing the sensor activation for chair occupancy. Samples are collected every second and '1' indicate that chair is occupied.

$$
\begin{aligned}
A &= ...1111111000000011111...\\
B &= ...1111111100000001111...\\
C &= ...1111111000011001111...
\end{aligned}
$$

Table 4.1: A sample of similarity measures for binary sequences.

| Measure | S(A,B) | Source by [141] |
|---|---|---|
| Jaccard | $\frac{M_{11}}{M_{11}+M_{01}+M_{10}}$ | Jaccard(1901) |
| Dice | $\frac{M_{11}}{2M_{11}+M_{01}+M_{10}}$ | Dice and Sorenson(1945) |
| Hamming | $N - M_{01} + M_{10}$ | Hamming(1950) |
| Faith | $\frac{M_{11}+0.5}{M_{11}+M_{01}+M_{10}+M_{00}}$ | Faith(1982) |
| Gower & Legendre | $\frac{M_{11}+M_{00}}{M_{11}+0.5(M_{01}+M_{10})+M_{00}}$ | Gower and Legendre(1971) |

Both sequences B and C differ from sequence A by a Hamming distance of two. However, intuitively B is closer to A than C to A. Sequence A shows that a person sat down on a chair initially and then for seven seconds stood-up before they sat down again. More or less the same pattern is repeated in sequence B with a one second delay in standing up and a one second delay in sitting down again. However, in sequence C, the user sat down for two seconds. Therefore, to measure not only the number of mismatches but also how close the local changes are (how far apart the mismatches happen) a fuzzy distance measure (fuzzy hamming distance) is required. A detailed explanation about the fuzzy distance measure (or also known as generalised Hamming distance) is given in [152].

### 4.4.2 Dynamic Time Warping

The Dynamic Time Warping (DTW) algorithm is based on the concept that the similarity between two sequences can be assessed by aligning important patterns [154] [155]. This can be achieved by locally deforming the time axis in order to minimise the cumulative distance between the aligned points. This method is suitable for matching time sequences containing patterns that are qualitatively similar but have different paces and lengths in time.

Consider a test sequence $X = (x_1, \ldots, x_i, \ldots x_N)$ and a reference sequence $Y = (y_1, \ldots, y_j, \ldots y_M)$. A correspondence between their elements is established through the warping curve $\Phi = (\phi_t, \psi_t)$, with $t = 1, 2, \ldots, T$. In DTW to compare between two different signals, the similarity between two binary time series are measured using the Euclidean distance [156]. A warping curve is then calculated, thus:

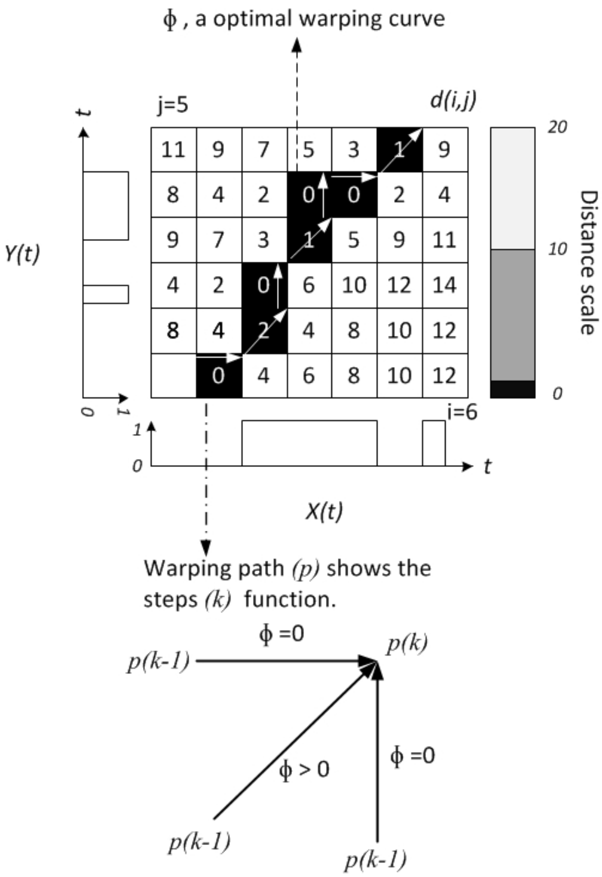



Figure 4.2: The optimal warping path aligning by two different binary signals.

$(d(x_{\phi_t}, y_{\psi_t}) = (X(i)-Y(j))^2$ where $j \in [1, M]$ and $i \in [1, N]$. With $i(k) \in [1, ..., I]$ and $j(k) \in [1, ..., J]$, where $k$ is step function, the warping path $(p)$ as shown in Figure 4.2 is restricted by several constraints as follows [156]:

- Continuity: Given $p_k(i(k), j(k))$ and $p_{k-1}(i(k-1), j(k-1))$, where $i(k) - i(k-1) \leq \mathbb{R}$ and $j(k) - j(k-1) \leq \mathbb{R}$ with $\mathbb{R}$ an integer value.

- Endpoint: The warping path will start at $i(1) = 1, j(1) = 1$ and stop at $i(k) = I, j(k) = J$.

- Monotonicity: Given $p_k(i(k), j(k))$ and $p_{k-1}(i(k-1), j(k-1))$, where $i(k)-$

$i(k-1) \geqslant 0$ and $j(k) - j(k-1) \geqslant 0$.

Among all the warping curves allowed, the optimal $\hat{\Phi}$ is taken which minimises the distance between the original and warped sequences. This is summarised as:

$$\hat{\Phi} = (\hat{\phi}_t, \hat{\Psi}_t) = \arg\min_{\phi_t, \psi_t} \sum_{t=1}^{T} \frac{d(x_{\phi_t}, y_{\psi_t}) m_{t,\Phi}}{M_\Phi} \tag{4.11}$$

where $m_{t,\Phi}$ is a local weighting coefficient and $M_\Phi$ is a path-dependent normalisation such that $M_\Phi = \sum_t m_{t,\Phi}$. Symbol $d$ is a local distance measure where its arguments, $x_{\phi_t}$ and $y_{\phi_t}$ are the elements of the warped input sequences. According to Pavel Senin [157] $d$ is the cumulative distance matrix ($c$) which is defined as follows:

- First row: $d(1,j) = \sum_{k=1}^{j} c(x_1, y_k), j \in [1, M]$
- First column: $d(i,1) = \sum_{k=1}^{i} c(x_k, y_1), i \in [1, N]$
- Others: $d(i,j) = min(d(i-1, j-1), d(i-1, j), d(i, j-1)) + c(x_i, y_j, i \in [1, N], j \in [1, M])$

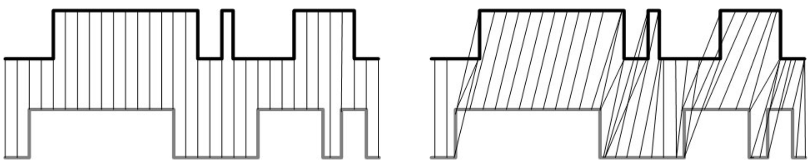
Given an ordered pair of sequences, the minimum cumulative distance is finally obtained along the optimum mapping $\hat{\Phi}$ [154]:

$$D_{DTW}(X,Y) = \sum_{t=1}^{T} \frac{d(x_{\hat{\phi}_t}, y_{\hat{\psi}_t}) m_{t,\hat{\Phi}}}{M_{\hat{\Phi}}} \tag{4.12}$$

Algorithm #1 shows how the optimum mapping $\hat{\Phi}$ could be found by the simple tracking from the first value to the last value of sequences $X$ and $Y$ [154] [157]. Figure 4.2 shows the optimal warping path aligning based on two different time series signals. A dark-light colour map is used to show the value of the distances.

**Algorithm #1: Optimal Warping Path**

```
path[]
i=rows(dtw)
j=columns(dtw)
while (i > 1) & (j > 1) do
    If i == 1 then
      j = j-1
      else if j == 1 then
      i = i -1
      else
      if dtw(i-1,j) ==min[dtw(i-1,j);dtw(i,j-1);dtw(i-1,j-1)] then
      i=i-1
      else if dtw(i,j-1) == min[dtw(i-1,j);dtw(i,j-1);dtw(i-1,j-1)] then
      j=j-1
      else
      i=i-1;j=j-1;
    end if
end while
return path
```



(a) Linear time alignment

(b) Nonlinear dynamic time warped

Figure 4.3: Linear and non-linear time alignment a) linear time alignment, where $i^{th}$ point in one sequence is aligned with $i^{th}$ point in other b) non-linear time warped alignment allows more distance alignment in similarity computation.

Similarity measures could be considered as linear time alignment. However, DTW is a non-linear time-warped alignment. In linear time alignment, the $i^{th}$ point in one sequence is aligned with the $i^{th}$ point in the other sequence. This is illustrated in Figure 4.3. The non-linear time warped alignment allows the generation of a better, more natural distance measure for use in similarity computation. Evidently, DTW is different from the others binary measures because of its non-linear adjustment to the similarity measure.

## 4.5 Discussion

This chapter gives an overview of different techniques to measure the distance or similarity between activities. The presented techniques are in two categories of statistical and distance measures. Application of consistency measure and ApEn measure in user activities recognition is novel and it has not be considered in the past. Some of the commonly used distance measures along with DTW techniques are briefly described in this chapter. These techniques are applied in ADL recognition.

Chapter 5 will describe application of the techniques presented in this chapter to the data presented in 3.

# Chapter 5

# User Profiling

## 5.1 Introduction

The analysis of the office workers' Activities of Daily Working (ADW) in an office environment can be used to optimize the energy consumption and also office workers' comfort. Therefore, it is essential to identify individual user profiles according to the user's preferences and behaviour. The individual profile will be used to automatically adjust office conditions according to the user preferences. Sensory signal outputs from monitoring system are used to recognise user activities and ultimately create a profile for each user. Statistical techniques and similarity measure techniques are employed to develop a simple user profile. Experimental results from real and simulated data are presented to validate the proposed methodologies for user profiling.

This chapter is organised as follows; in Section 5.2 user activity characterisation is explained. In Section 6.2 the implementation of conversion for collected data into meaningful format is presented. The validation of ADW data from user's behaviour in an office environment, and constructing the simple user profile are presented in Section 5.4. Details of experiments for measuring consistency and chaotic nature of user's activity are presented in Section 5.6 followed by experiments of thermal comfort monitoring in Section 5.7. The experimental results measuring similarities of behavioural pattern are presented in Section 5.8, and then the discussion is provided in Section 5.9.

## 5.2 User Activity Characterisation

User activity characteristics are those traits that differentiate a user from other users or the same user in two different time periods. For example coming to work early or working late evening are characteristics of a specific user different from other users. In our first attempt to construct individual user profile based on the information collected from the office environment, it is essential to characterise user activities. User activities and behaviour characterisation will lead to a user profile. The user profile then is used to interact with the environment and controlling different elements if necessary. An individual user profile is associated with the user's work habit and his/her preference.

To be able to quantify user characteristics, a data collection system is used to monitor appropriate behavioural and environmental conditions. A schematic diagram of the proposed methodology for data collection and user profiling is shown in Figure 5.1. Nine different types of sensors are suggested to be employed to monitor user's behaviour and thermal comfort in the office environment. Sensors are in two groups of ambient office monitoring and user's behaviour monitoring. User's behaviour represents the work habit and the ambient office monitoring

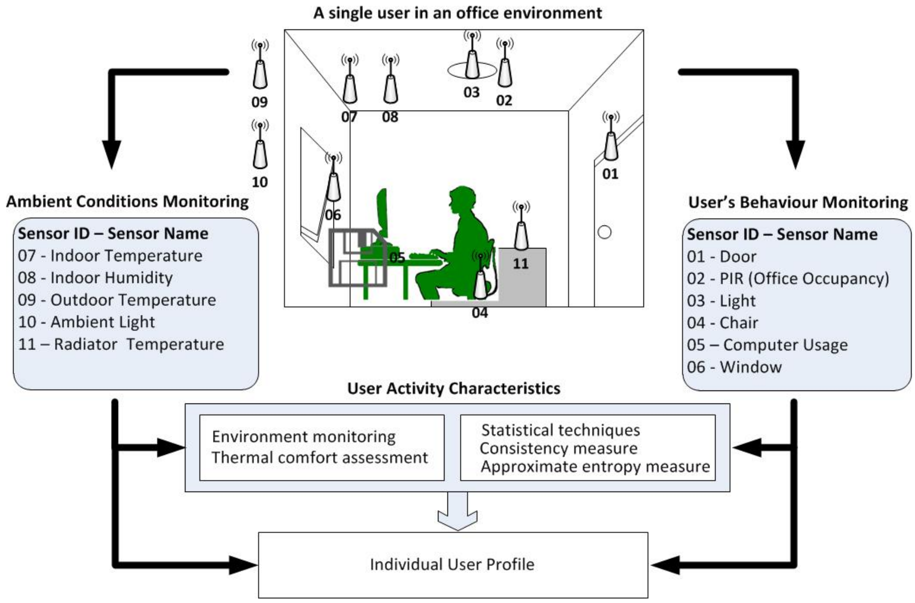


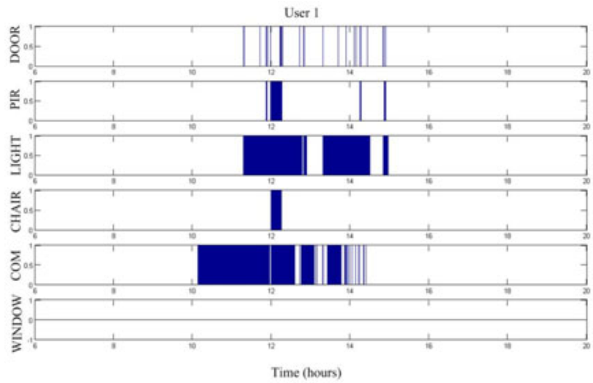

Figure 5.1: The proposed methodology to construct an individual user profile.

sensors represent the thermal comfort and to a certain degree the energy consumption. Gathered information from the office environment are represented in two main groups; ambient conditions and user's behaviour monitoring.

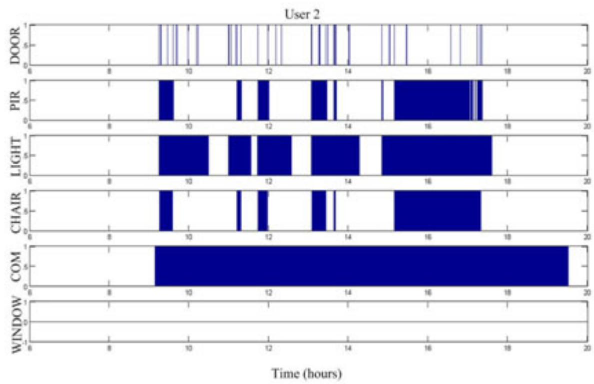
To characterise a user's behaviour, the following main categories are identified.

- **Working Time** depends on the user's arrival and departure. The information is important because it can be used to classify office user attitude such as early starter or late starter. This information can be used to control heating, lighting system and computer power, so the office's power can be turned on in anticipation to the users arrival and turned off on departure.

- **Office Power Usage** is based on the measured duration time of office's electrical devices (e.g. computer,lighting and heating) that are used by user. This information is important to classify office user working attitude such as light, medium and heavy user of office's electrical devices.




(a) User #1




(b) User #2

Figure 5.2: A sample of daily pattern for a) User #1 b) User #2.

- **Work Routine** is associated with consistency or how long a user maintains the behaviour during working in an office environment. A work routine is predictable. This information is important to classify office user working attitude such as very routine, routine and random.

- **User's Thermal Comfort** is the environmental factors such as humidity and source of heat in the workplace combined with the personal factors such as clothing and how physically demanding the work is.

## 5.3 Data Representation

The development of data collection system is explained earlier in Chapter 3. Upon collection of the data from different sources, it is important to find ways of representing the data in a meaningful manner. The raw sensory data is often difficult to visualise and understand. This will be even more complicated when sensors from different sources and different types are gathered. For example, consider a sample of activities for two different users; User #1 and User #2 shown in Figures 5.2-a and 5.2-b respectively. Presented data are from 6 sensors in binary format only. It is important to summarise the collected data before it is used to construct a user profile. So, to summarise collected data into meaningful knowledge, investigated techniques are explained below.

### 5.3.1 Representation of Binary Signals in Start-time and Duration Sequences

To visualise a long sequence of collected data, different forms of data visualisation are considered. To achieve this, raw data sequences can be converted into start-time and duration sequences. This will reduce the dimensionality of the stored data and it will also help to visualise a specific activity for a longer period of time. Consider a binary sequence, $s(t)$, representing the office door opening for $t = 1, 2, \ldots, N$, where $s(t) \in [0\,1]$. The signal has two states of 'open' and 'close' representing the status of door.

$$s(t) = (\ldots, 0, 1, 0, 0, \ldots, 0, 0, 1, 0, 0, \ldots) \tag{5.1}$$

To represent the signal in start-time and duration, initially the signal is converted into start-time and stop-time, $x(t)$, as stated below:

$$x(t) = (t_{s_1}, t_{e_1}, t_{s_2}, t_{e_2}, \ldots, t_{s_i}, t_{e_i}, \ldots, t_{s_n}, t_{e_n}) \tag{5.2}$$

where $t_{s_i}$ and $t_{e_i}$ are the start-time and stop-time of any entry with a value of 1 in $s(t)$. Therefore, start-time sequences, $ST(t)$ and duration sequences $DU(t)$ are represented respectively as:

$$ST(t) = (t_{s_1}, t_{s_2}, \ldots, t_{s_i}, \ldots, t_{s_n}) \tag{5.3}$$

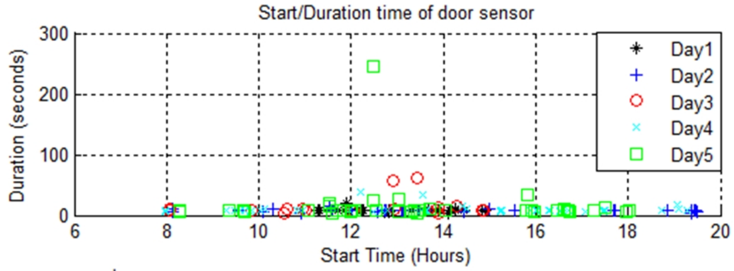
$$DU(t) = (t_{e_1} - t_{s_1}, t_{e_2} - t_{s_2}, \ldots, t_{e_i} - t_{s_i}, \ldots, t_{e_n} - t_{s_n}) \tag{5.4}$$

For example, using start-time and duration form of representation, in Figures 5.3-a and 5.3-b door activities over 5 working days are illustrated for User #1 and User #2 respectively.

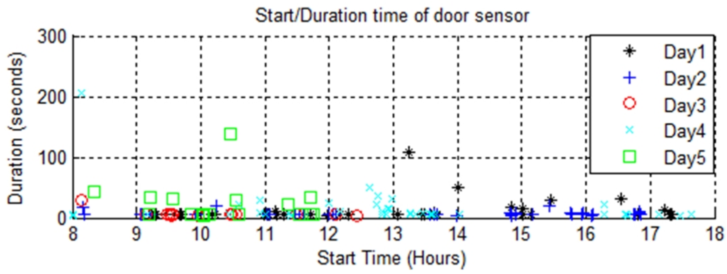
### 5.3.2 Representation of Binary Signals in Binary Code

The sensor output for door, PIR, chair and window are represented in binary format. Therefore it is possible to combine all and represent all these binary sequences in one sequence with real values. To achieve this, the binary bit representing the state of a signal at any instance is merged together to generate an $n$ bits binary code where $n$ represents the number of signals.

For example if $s_1(t)$ represents chair occupancy and $s_2(t)$ represents windows opening;




(a) User #1




(b) User #2

Figure 5.3: The door activities over five working days for a) User #1 b) User #2.

$$s_1(t) = (\ldots, 0, 1, 0, 0, \ldots, 0, 0, 1, 1, 1, \ldots) \quad (5.5)$$

$$s_2(t) = (\ldots, 0, 1, 1, 1, \ldots, 0, 0, 1, 0, 0, \ldots) \quad (5.6)$$

then combined bits binary code will be:

$$x(t) = (\ldots, 0, 3, 1, 1, \ldots, 0, 0, 3, 2, 2, \ldots) \quad (5.7)$$

There is no basis here as to which signal should be on lower or higher bit. Therefore, the above combination could also be represented as:

$$x(t) = (\ldots, 0, 3, 2, 2, \ldots, 0, 0, 3, 1, 1, \ldots) \quad (5.8)$$

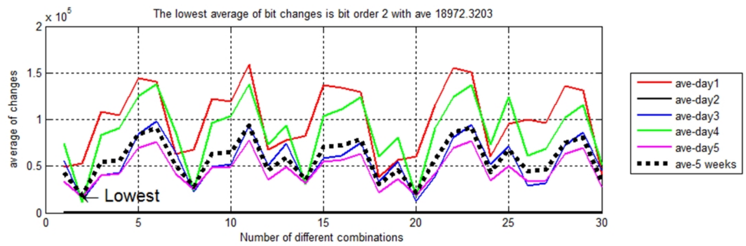
Based on some initial results and investigation, it is decided to have a weighted approach to integrate the binary signal. Signals will be integrated based on the following expression:

$$\begin{aligned} x(t_i) = \quad & (S1_{t_i} \times W_1) + (S2_{t_i} \times W_2) + (S3_{t_i} \times W_3) + (S4_{t_i} \times W_4) + \\ & (S5_{t_i} \times W_5) + (S6_{t_i} \times W_6) \end{aligned} \tag{5.9}$$

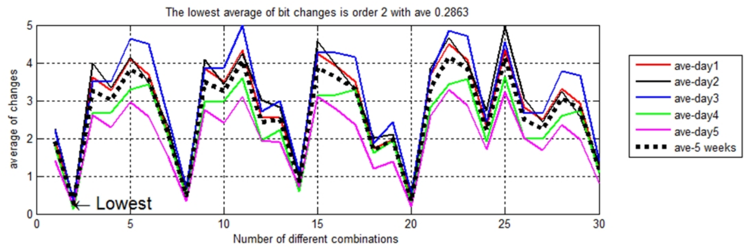
where $t$ is time of unified signal, $i$ sequence number of data based on real time, $S1, S2, S3, S4, S5, S6$ are sensory signals in bit order, and $W_i$ is weight values bit. Weight values of $W_i$ are the most significant bit (MSB) is the bit in a multiple-bit binary number with the largest value, while the least significant bit with smallest value.

To identify order of bits, it is decided to check bit changes from one state to the next as a representative of the important bits. As the number of signal increases, it will become harder to identify the order of bits. It is proposed to use Hamming distance to measure the distance as a representative of changes.

For the daily patterns of User #1 and User #2 shown earlier, $n = 6$ bits binary code is generated and all possible combination of sensors are tested (30 combinations). Figure 5.4-a shows for User #1 the combination sensors of lowest average of bit changes is achieved from the 2nd combination. Figure 5.4-b shows signal combination for User #2 and again the lowest average of bit changes is also 2nd combination. Therefore, it is concluded that 2nd combination i.e.




(a) User #1



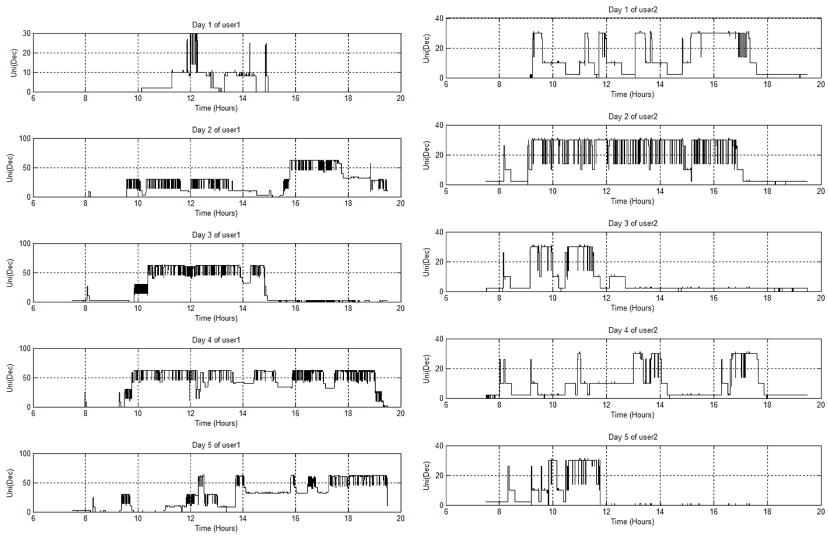

(b) User #2

Figure 5.4: Distance representing bit changes for different combination of signals a) user #1 and a) user #2.

$$[2^5 = window, 2^4 = door, 2^3 = PIR, 2^2 = light, 2^1 = chair, 2^0 = PC]$$

is producing less changes and as a result a preferred choice for combining binary signals.

Using the signal integration mentioned in this section, the combined signal for the samples of activities of two users are shown in Figure 5.5. From figure 5.5-a, it can be observed that User #1 had less activity on day 1 and did not use office on day 3 after 3:00 PM. For User #2, in Figure 5.5-b it is observed that user is busy on day 2 and did not use office on day 5 after 12:00 PM.



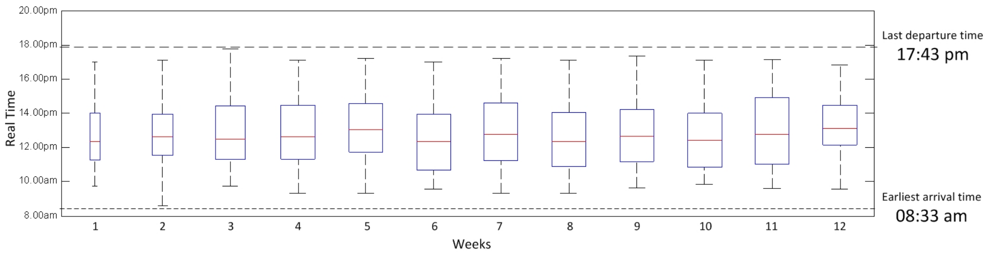

(a) User #1 (b) User #2

Figure 5.5: Combined pattern of activities a) user #1 and a) user #2.

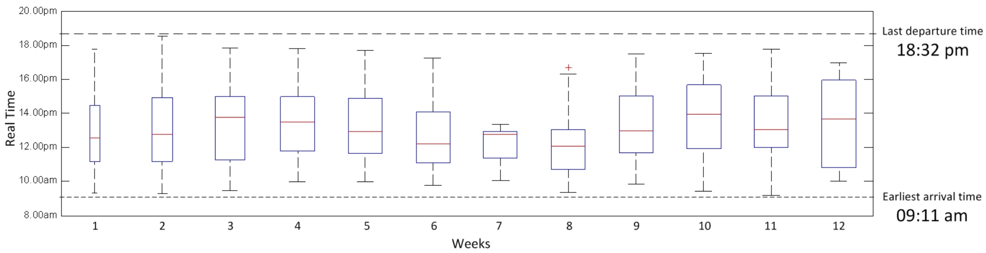
## 5.4 Annotation and Data Validation

Validation or verification are methods to examine that the data collected by data collection system is correct. The examination of behaviour pattern of users also means that inspect the performance of a data collection system to recognize the user activity correctly. Sometimes, abnormal activity is recorded caused by faults of data collection system (e.g. a sensor node is run out of battery) or some of the activity is totally out of routine (e.g. change of room layout or furniture).

For example, statistical results of office occupancy and chair occupancy in Table 1, Table 2 and Table 3 (in Appendix D), that there are showed normal activity of user's activities. From the week 1 to 12, most of the durations of chair occupancy are less than the durations of office occupancy. Detail annotation activity and validation of single user's activity in an office environment are presented in Appendix F.




(a) User #1



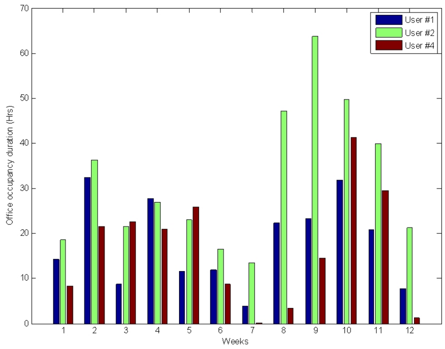

(b) User #4

Figure 5.6: Distribution of door activity over twelve weeks for a) User #1, b) User #4.

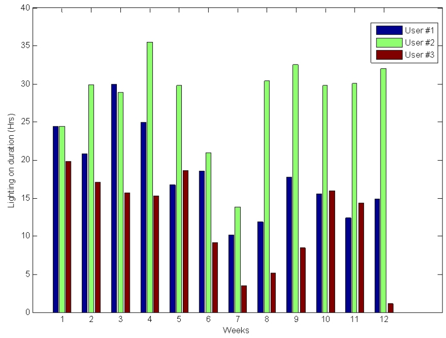
## 5.5 Statistical Measures

Signals collected from different sensors will vary for different users and they also vary across different working days. To be able to understand and also summarise different uses' activities, after the validation of the data, it is necessary to carry out statistical analysis of the collected data.

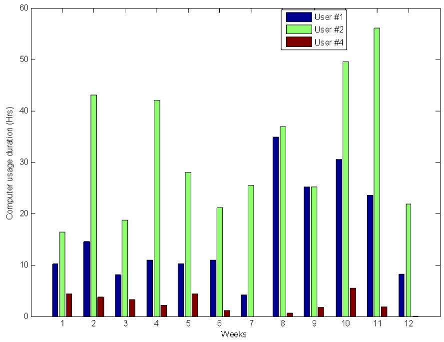
To depict groups of numerical data through their statistical quantities, a Box Plot is a convenient way of illustrating information graphically. The range of the vertical scale is from minimum to the maximum values. Each vertical scale also hold information about the lower quartile ($Q_1$), mean or median, and upper quartile ($Q_3$). Interquartile range (IQR) is the difference between $75^{th}$ percentiles, $Q_3$, and $25^{th}$ percentiles, $Q_1$, of distribution data. Figures 5.6-a and 5.6-b show distribution of office door activities over twelve weeks for User #1 and User #4



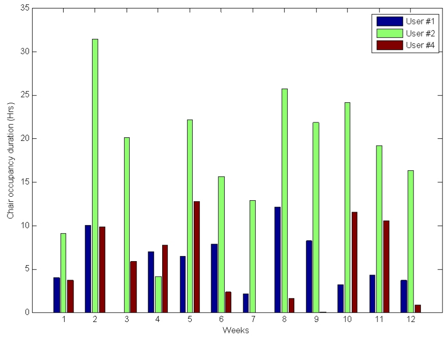

(a) Office occupancy




(b) Lighting usage




(c) Computer usage




(d) Chair occupancy

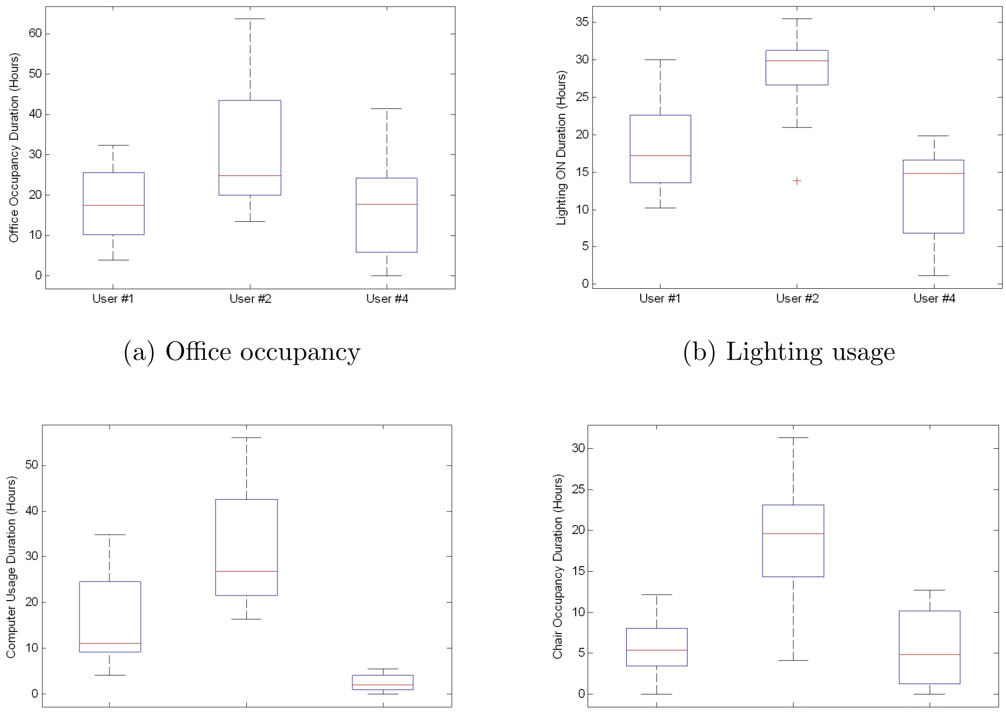
Figure 5.7: Comparative results of users behaviours based on weekly duration of a) Office occupancy, b) Lighting usage, c) Computer usage, d) Chair occupancy.

respectively. From these figures, it could be observed that users **normally** arrive between 08.00 AM to 09.30 AM and leave the office **normally** between 05.00 PM to 06.30 PM. The patterns of arrival and departure times fluctuate, this may be caused by the schedules of work.

Bar charts are used to compare between different user's behaviour based on the measurement of the weekly duration of office occupancy, lighting usage, chair occupancy and computer activities. Figure 5.7 illustrate the comparative results between User #1, User #2 and User #4. The bars in graphs provide a visual display for comparing weekly duration for different user. In Figure 5.7-a, the weekly durations' values show that User #2 spent more time in the office followed by User #1 and User #4. From Figures 5.7-b and 5.7-c it can be observed that more lighting and computer usage time are recorded for User #2 followed by User #1 and User #4. These results highlight the facts that each user has dissimilar ADW .

In order to construct an individual user profile, a box plot is also used to extract information from the user behavioural data. The distribution of weekly duration of the user activities using box plot is shown in Figure 5.8. The boxes provide a visual display for comparing duration of different activities of users. Comparing with bar charts in Figure 5.7, it can be seen that the box plot described more details about the dissimilar behaviour based on weekly durations of user activities. From Figure 5.8, it can be observed that User #2 spent more time in office and seating in the chair compared to other users. User #2 has also used computer and room lighting more than any other users over twelve weeks. It is also possible to identify any anomaly within the distributed data. For example in Figure 5.8-b it is shown that anomalous activity for lighting of User #2 is detected.

Full details of the statistical parameters calculated for the experimental data are presented in Appendix D. Table 1, Table 2 and Table 3 in Appendix D show summaries of weekly statistical analysis results of sample data from User #1, User #2 and User #4.




(a) Office occupancy (b) Lighting usage

(c) Computer usage (d) Chair occupancy

Figure 5.8: Box plot representing users behaviours based on weekly duration of a) Office occupancy, b) Lighting usage, c) Computer usage, and d) Chair occupancy.

## 5.6 Consistency and Chaotic Measures of User's Activity

The consistency of user's activity is measured using Cronbach's Alpha ($\alpha$) explained in Section 4.3.1. Chaotic measure in user's behaviour is also measured using Approximate Entropy (ApEn) explained in Section 4.3.2. To illustrate the effect of measures for different user's activities, data set $D4$, $D5$ and $D6$ in Table 3.2 are used for our investigation. The daily duration of user's activity is used to investigate the consistency level, and daily number of changes of user's activity is used to investigate the chaotic levels of activities.

### 5.6.1 Consistency Measure for User Behaviour

To apply Cronbach‘s Alpha ($\alpha$) technique to measure internal consistency of data, several parameters of user's activities such as office occupancy, lighting activity, chair occupancy and computer activity are selected. Initially, data are presented in a matrix where rows represent changes in a week, and columns present daily duration of activities. For our data collected over twelve weeks (5 working days per week), a matrix $12 \times 5$ is created.

Applying the mentioned calculation process explained in Section 4.3.1, user's activities data of office occupancy, lighting, chair and computer are rescaled and degrees of level of ($\alpha$) values are represented as internal consistency of activities. The objectives of measure the activity's consistency using Cronbach's Alpha are to determine the level of consistency based on range 0 to 1, and then distinguish the overall behaviour between users. The value of $\alpha$ is considered reasonable consistency at $\alpha > 0.5$, while alpha values above 0.7 indicate extremely high consistency [158].

Table 5.1: Cronbach‘s Alpha ($\alpha$) values of user's activity in an office environment

| | Office Occupancy | Lighting activity | Chair Occupancy | Computer activity |
|---|---|---|---|---|
| User #1 | 0.5134 | 0.5129 | 0.5973 | 0.4619 |
| User #2 | 0.4337 | 0.1727 | 0.4337 | 0.3453 |
| User #4 | 0.6661 | 0.6092 | 0.7631 | 0.6092 |

Table 5.1 presents the internal consistency of user's activities in an office environment. From the results, it can observed that User #4 shows consistency with all activities at the degree level of $\alpha > 0.5$, while chair occupancy of User

#4 is considered extremely high consistency with $\alpha > 0.7$. Behaviour of User #1 is almost considered reasonable of consistency, where office occupancy, lighting activity and chair occupancy are at $\alpha > 0.5$, but computer activity is below satisfaction of the consistency level. While, User #2 activities is have consistencies of all ADW in an office environment at $\alpha < 0.5$. The results can be regarded as simply to compare and allow for extraction of the user's profile based on the activity consistency level in an office environment.

Table 5.2: ApEn(2,r,N) calculations for User #1, User #2 and User #4

| | ApEn(2,r,N), where r=0.1*SD and N=84 days | | | | | |
|---|---|---|---|---|---|---|
| | User #1 | | User #2 | | User #4 | |
| Activity | SD | ApEn | SD | ApEn | SD | ApEn |
| Office Occupancy | 10.1323 | 0.6904 | 9.7696 | 0.5612 | 6.3148 | 0.6631 |
| Lighting | 2.4856 | 0.7990 | 4.1000 | 0.5494 | 3.0464 | 0.7856 |
| Chair | 5.8131 | 0.6231 | 4.9141 | 0.6746 | 3.9591 | 0.7106 |
| Computer | 3.8967 | 0.6860 | 7.4720 | 0.4476 | 0.4330 | 0.3288 |

| | ApEn(2,r,N), where r=0.2*SD and N=84 days | | | | | |
|---|---|---|---|---|---|---|
| | User #1 | | User #2 | | User #4 | |
| Activity | SD | ApEn | SD | ApEn | SD | ApEn |
| Office Occupancy | 10.1323 | 0.7403 | 9.7696 | 0.8283 | 6.3148 | 0.7805 |
| Lighting | 2.4856 | 0.7990 | 4.1000 | 0.5494 | 3.0464 | 0.7856 |
| Chair | 5.8131 | 0.7743 | 4.9141 | 0.6746 | 3.9591 | 0.7106 |
| Computer | 3.8967 | 0.6860 | 7.4720 | 0.6017 | 0.4330 | 0.3288 |

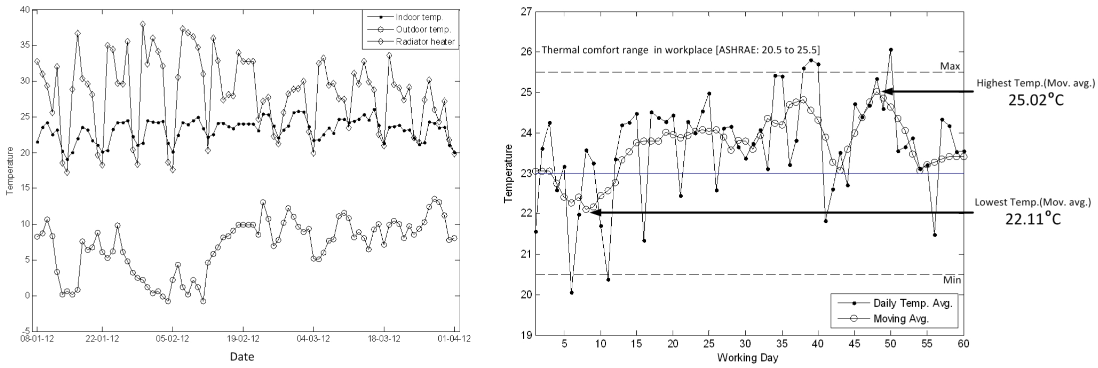
### 5.6.2 Approximate Entropy Measure for User Behaviour

The ApEn is a measure that depends on a sequence of daily number of changes from users behaviours for 84 days (12 weeks) period. Several parameters from the user's activities such as office occupancy, lighting activity, chair occupancy and computer activity are selected for analysis. According to a study in [150], ApEn require $m$ parameter as embedding dimension and $r$ parameter as a tolerance value. Therefore in this experiment $m = 2$ is selected and $r$ is used based on 0.1 and 0.2 times standard deviation of data. Selected values follow the recommendation by Sheng Lu et al [159]. Then, the calculation process discussed in Section 4.3.2 is applied.

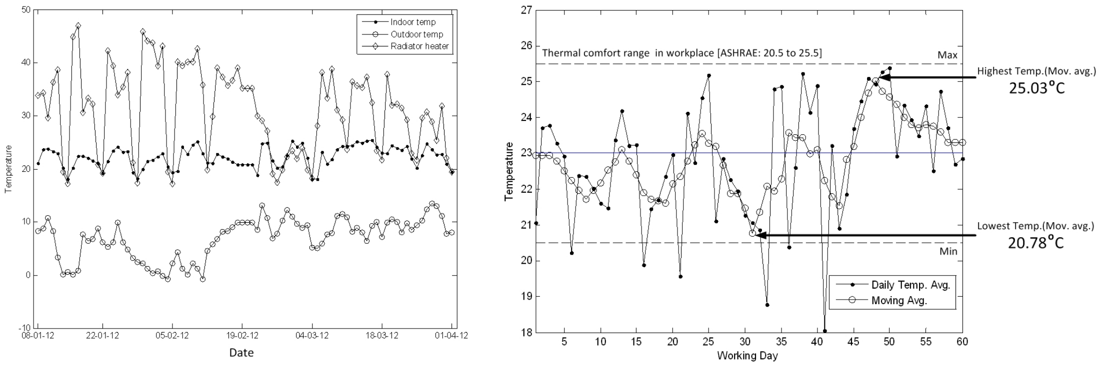
Table 5.2 shows the calculations of ApEn for User #1, User #2 and User #4. The values of ApEn can be used to compare the behaviours of users based on chaots level. According to Kalon et al. [160], very small values of ApEn's implies the sequence is less chaotic and predictable. A study by Sheng Lu et al [159] tested three signals. The results show that ApEn values are 1.157 for white noise signal and 0.637 for cross chirp signal and 0.015 for the sinusoidal signal. Therefore, based on these the user's behaviours are compared based on ApEn values. We found that all chaotic levels of users behaviours are acceptable. Moreover, User #2 behaviour's signals are less chaotic (ApEn less than 0.7 at $r = 0.1 * SD$), however at $r = 0.2 * SD$ office occupancy shows increased with ApEn more than 0.8. While, User #1 and User #4 show their behaviour for all activities are below ApEn values of 0.8. Computer activity of User #4 with ApEn less than 0.4 shows this to be more predictable compared to her/his other activity or other users activities.

## 5.7 Thermal Comfort Monitoring and Assessment

To assess the thermal comfort in the office environment, ambient room temperature, outside temperature and radiator temperatures are measured. It is expected that these measurements will provide us with the thermal comfort range. For our investigation, data sets D4, D5 and D6 are used.




(a) User #2




(b) User #4

Figure 5.9: The office temperature readings for a) User #2 b) User #4.

Figure 5.9-a and Figure 5.9-b show the heating temperature from radiator heater in an office is significant due to the low temperature of outdoor temperature for User #2 and User #4 respectively. Since a radiator heater is ON, it will change the indoor thermal comfort. The arrow lines are used to locate points on maximum and minimum values of temperature moving average to define the thermal comfort range of user's office. The moving average is computed by averaging with five data points as five working days (Monday to Friday) in a week of data over sixty days. Full results of statistical analysis of the room temperature for the above data sets are presented in Table 4 in Appendix D.

The temperature ranges to meet the thermal comfort satisfaction during winter months is between 20.5 °C to 25.5 °C based on suggested range by Canadian Centre for Occupational health and Safety [39].

## 5.8 Measuring Similarities of Behavioural Patterns

To compare behavioural patterns, similarity measures are used. Both linear techniques and non-linear measure are considered in this investigation. In this investigation, initially data pre-processing of temporal sequence data from sensory signals is conducted. Different similarity measure are tested and the results are used to analyse user's behavioural patterns. Well-known linear similarity measures including Dice, Faith, Jaccard (JCD), Gower and Legendre (GL), and Hamming Distance (HD) are used. For non-linear similarity measures, Dynamic Time Warping (DTW) is used. Presented results in this section are based on the experimental data sets $D4$, $D5$ and $D6$.

Considering the fact that the offices used for this investigation are academic offices, pattern of usage has varied during the academic semester. During semester breaks, user's behaviour would change; which meant they would either be in an office for a longer time or absent. Therefore, similarity measures are used in this study to identify the similarity of user's behaviour in an office environment, whereby we could identify user's behaviour of office use over a period of time.

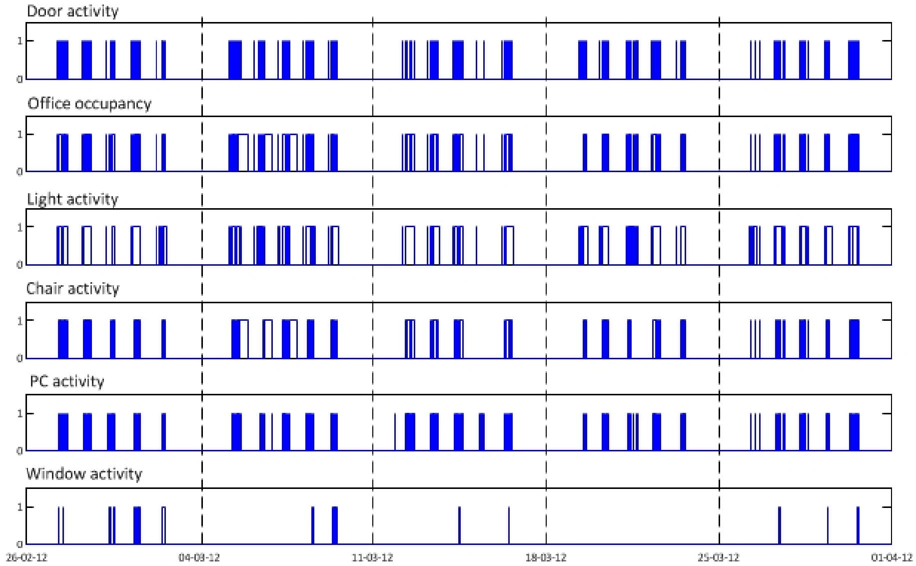



Figure 5.10: The behavioural pattern of User #2 in an office environment over five weeks.

### 5.8.1 Linear Similarity Measures

In order to show the difference in performances of conventional similarity measures, binary similarity using the named techniques are measured for $D5$ data set (User #2). Figure 5.10 represents the behaviour of user in binary data format for over five weeks. Based on these binary signals, Table 5.3 shows a summary of the similarity values for conventional linear similarity measures. The similarities of five working days are compared against day 1 of User #2. This is repeated for all similarity measures.

Similarity measures Dice and JCD show similar performances. GL method showed a strong similarity between the days of the week; whilst the other methods showed conflicting results. Faith distance similarity provides a much more consistent degree of similarity than Dice, JCD and GL. However, HD gives a number of mismatching bits between two binary sequences of the same length. When

Table 5.3: The summarised results of conventional similarity measures

| Data set (43200 samples) | | Day 1 of User #2 | | | | |
|---|---|---|---|---|---|---|
| Measures | Day | Dice | Faith | JCD | GL | HD |
| Door | 1 | 1 | 1 | 1 | 1 | 0 |
| | 2 | 0.07 | 0.5 | 0.03 | 0.99 | 271 |
| | 3 | 0 | 0.49 | 0 | 0.99 | 280 |
| | 4 | 0.83 | 0.5 | 0.71 | 1 | 274 |
| | 5 | 0 | 0.5 | 0 | 0.99 | 280 |
| PIR | 1 | 1 | 1 | 1 | 1 | 0 |
| | 2 | 0.81 | 0.64 | 0.68 | 0.86 | 8940 |
| | 3 | 0.75 | 0.57 | 0.61 | 0.84 | 16601 |
| | 4 | 0.99 | 0.85 | 0.99 | 1 | 6601 |
| | 5 | 0.61 | 0.46 | 0.44 | 0.76 | 14504 |
| Light | 1 | 1 | 1 | 1 | 1 | 0 |
| | 2 | 0.86 | 0.72 | 0.76 | 0.89 | 8035 |
| | 3 | 0.84 | 0.67 | 0.73 | 0.89 | 21911 |
| | 4 | 1 | 0.91 | 0.99 | 1 | 5919 |
| | 5 | 0.68 | 0.51 | 0.51 | 0.74 | 14260 |
| Chair | 1 | 1 | 1 | 1 | 1 | 0 |
| | 2 | 0.8 | 0.61 | 0.66 | 0.86 | 8685 |
| | 3 | 0.68 | 0.51 | 0.51 | 0.81 | 16634 |
| | 4 | 0.99 | 0.84 | 0.99 | 1 | 6573 |
| | 5 | 0.59 | 0.45 | 0.42 | 0.76 | 13489 |
| Computer | 1 | 1 | 1 | 1 | 1 | 0 |
| | 2 | 0.35 | 0.45 | 0.21 | 0.92 | 4886 |
| | 3 | 0.26 | 0.44 | 0.15 | 0.93 | 5325 |
| | 4 | 0.99 | 0.52 | 0.99 | 1 | 4822 |
| | 5 | 0.16 | 0.39 | 0.09 | 0.86 | 5245 |
| Window | 1 | 1 | 1 | 1 | 1 | 0 |
| | 2 | 0 | 0.43 | 0 | 0.92 | 6333 |
| | 3 | 0 | 0.44 | 0 | 0.94 | 6334 |
| | 4 | 0.99 | 0.59 | 0.99 | 1 | 2040 |
| | 5 | 0.17 | 0.29 | 0.09 | 0.7 | 4320 |

comparing Faith and HD, there seems to be a relationship between their degrees of similarity results. From these observations, it is concluded that Faith and HD work best for all binary signals used to measure the similarity of activities.

It can be argued that Faith and HD linear measures represent similarities between activities best. Activities with the highest similarity between days or weeks indicate that the two sequences of activities are similar. In order to compare similarities between two activities on different days or weeks, another similarity test is conducted. In this similarity test, only HD is used to measure the degree of similarity for user activities in an office environment. Although the results shown in Table 5.3 demonstrated that Faith and HD are the best linear measures representing similarities between activities, only HD is used because it is able to separate between two different sequences using the number of mismatching bits; whereas using the Faith method only, used the range 0 to 1.

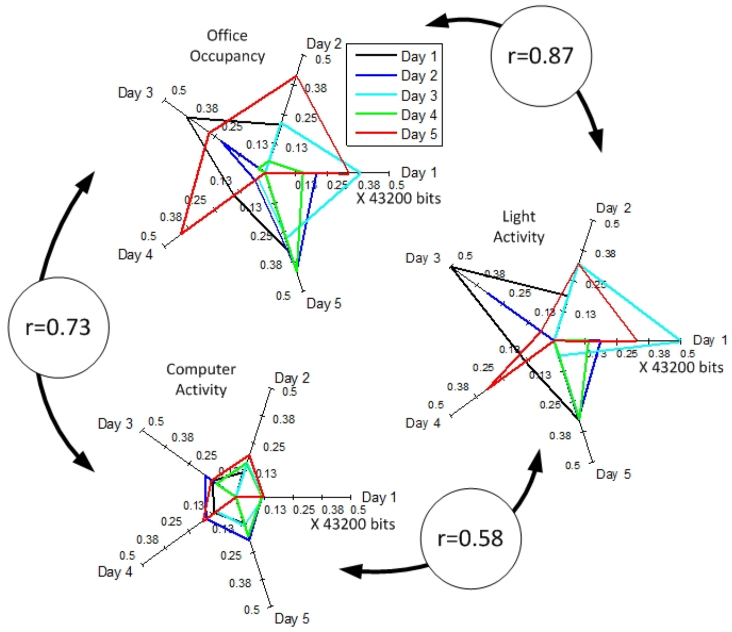
A calculation method to identify activities with the highest similarity between days or weeks is proposed as follows:

- Measure the similarity: $s(A_i, B_j)$
- Compute the total number of mismatching bits between all sequences (comparing activity on different days or weeks): : $S_T = \sum_{n=1}^{N} s_n(A_i, B_j)$, where $n$ is number of days or weeks and N is the total number of days or weeks.
- Find the minimum total number of mismatching bits for each day or week: $min(S_T)$.

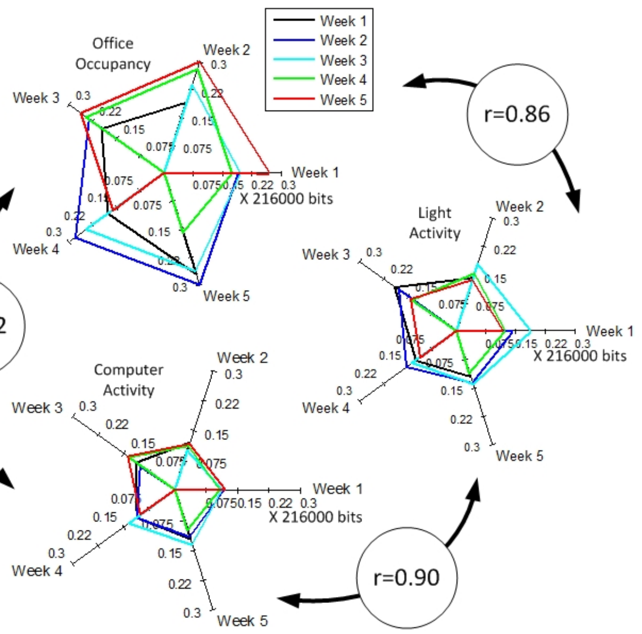
Based on the proposed method, the activities of User #2, such as office occupancy, chair activity, and computer activity, are measured using HD. Figure 5.11 shows the similarity measure results of office occupancy, light activity, and computer activity, for User #2 over five days as well as five weeks. Hamming similarity measure of user's behaviour between days over a period of five days, and then between weeks over a period of five weeks are computed. The spider plots in Figure 5.11 are plotted based on the similarity measure results as shown in Table 1 in Appendix E.

According to Table 1 in Appendix F, the bold numbers represent the highest similarity between two different sequences of activities. $min(S_T)$ is used to find the minimum total number of mismatching bits of HD for comparing between days and weeks. Within the similarity measurement results for a period of five days, the fourth day of office occupancy showed the highest similarity with $S_T =$ 28031 mismatching bits, the fourth day of light activity also showed the highest similarity with $S_T = 19888$ mismatching bits, while the first day of computer activity had the highest similarity with $S_T = 20280$ mismatching bits. For the similarity measure of User #2 activities for a period of five weeks, the first week for office occupancy showed the highest similarity ST = 179966 mismatching bits, the fifth day of light activity showed the highest similarity ($S_T = 109786$ mismatching bits), while the second week of computer activity had the highest similarity with $S_T = 96025$ mismatching bits. Comparing the results of similarity measures between days and weeks of activities for User #2, it can be concluded that any significant rise in the number of bits in the activities sequence gave rise to a corresponding increase in the number of mismatching bits of hamming distance.

In order to analyse user behaviour, spider plots are used to illustrate such similarities between days and weeks; and to investigate the relationship between activities based on the matrices of similarity scores. In Figure 5.11, the spider plots of HD scores showed that the similarity between weeks had nearly identical plots between activities for User #2. Therefore, Pearson's correlation coefficient, $r$, is used to test the degree of relationship between the activities of the user. The correlation coefficient, $r$, between two different similarity matrices ($A$ and $B$) is




(a)




(b)

Figure 5.11: Comparison of similarity measures for the activities of User #2 between a) 43200 samples per day for a period of 5 days, and b) 216000 samples per week for a period of 5 weeks.

computed using the following expression.

$$r = \frac{\sum_m \sum_n \left(A_{mn} - \bar{A}\right)\left(B_{mn} - \bar{B}\right)}{\sqrt{\left(\sum_m \sum_n \left(A_{mn} - \bar{A}\right)^2\right)\left(\sum_m \sum_n \left(B_{mn} - \bar{B}\right)^2\right)}} \tag{5.10}$$

where $\bar{A} = mean(A)$ and $\bar{B} = mean(B)$. A correlation coefficient less than 0.5 is representing a weak relationship and close to 1 is representing a highly correlated relationship [161].

In Figure 5.11, the correlation of similarity matrices for days comparison between office occupancy and light activity are high with $r = +0.87$, $r = +0.73$ for office occupancy and computer activity and $r = +0.58$ for between light activity and computer activities. Based on these correlation results, it can be concluded that computer usage and light usage are highly correlated to office occupancy. Meanwhile, computer usage and light usage showed a medium relationship. In contrast, the correlation of similarity matrices for weeks comparison showed a high relationship between computer usage and light usage with $r = +0.90$. However, the degree of relationship between office occupancy and light activity ($r = +0.86$) and computer activity ($r = +0.92$) for weeks comparisons are high and similar with the days comparisons.

Measure of correlation between two different similarity matrices giving a value between +1 and -1 inclusive. In this study, the correlation coefficient is divided into three levels of "low", "medium", and "high", as used in Figure 5.12. In the relationship model, these three levels of $r$ were used to define the strength of the linear relationship between two different user's activities in an office environment.

In order to model individual user behaviour for office energy, the results shown

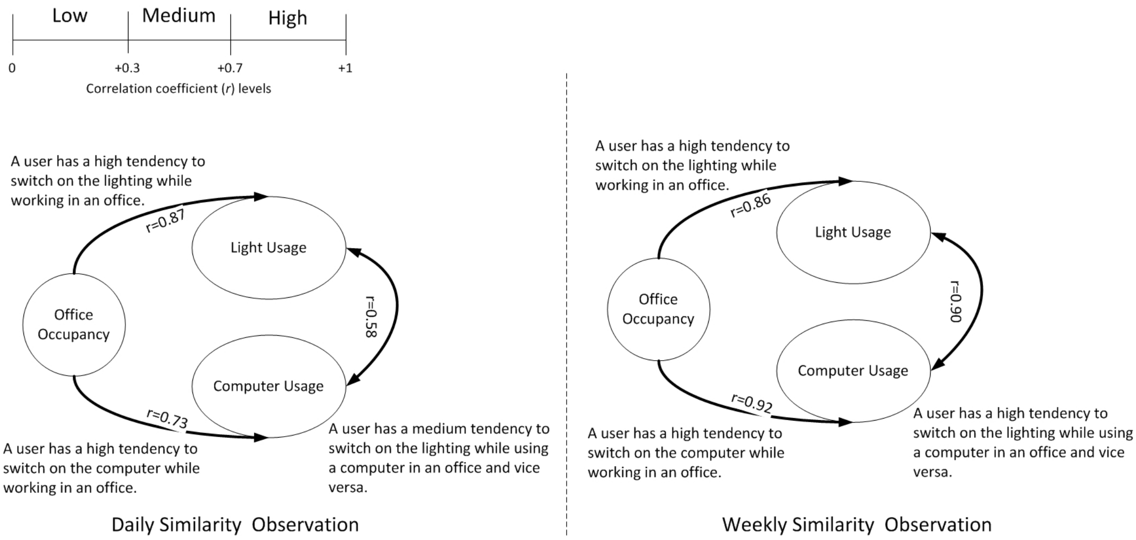



Figure 5.12: Energy use behaviour of User #2 in an office environment.

here are used to produce a relationship model of user behaviour. User activities, such as office occupancy, light activity, and computer activity, are considered as the Energy Use Activity (EUA) of a user. Figure 5.12 shows the relationship model for obtaining estimates of user's behaviour, based on energy consumption. This model could be used to provide information to an office environment control system, in order to learn the user's behaviour on energy consumption, while working in an office environment. For example, the environmental control system can be improved to understand of user's behaviour in determining user's tendency to move from an activity to another activity.

### 5.8.2 Similarity Features and Clustering

In order to identify the similarity between users' activities on different days, homogeneous sequence of activities are clustered together. Initially, daily activity sequences from a user is divided with a comparator sequence. A comparator sequence with the same length as the user's activity sequence and consists of logic 1 from 8.00 AM to 17.00 PM, and the rest is set to logic 0.

A pairwise representation method is proposed in [162], where similarity scores are used as a pairwise representation for visualizing and understanding purposes. This process is used to generate similarity scores for a pairwise representation of similarities. The polar dendrogram is used to show clustered data into a group of similar samples, thus producing a data structure using hierarchical clustering [162, 163]. In dendrogram clustering, the distance between two data points is computed by linkage creation. The number of clusters is chosen based on a distance rate of below 0.2.

To following steps are recommended to measure and visualise pairwise similarity values of user's activities:

1. Standardise the pairwise similarity values to range 0 to 1 the total bit sequence.

2. The dendrogram plot is used to cluster the pairwise values.

3. Normalise the standardised data of pairwise similarity values by dividing each data column (i.e., pairwise similarity values of office occupancy, light activity, chair activity, and computer activity are put in columns) by its standard deviation.

4. Use Principal Component Analysis (PCA) to generate the principal components values to visualise similarity data in different dimensions [164].

5. Use biplot to identify which variables (user's activities) contribute to the same direction in the plot of two principal components [165, 166].

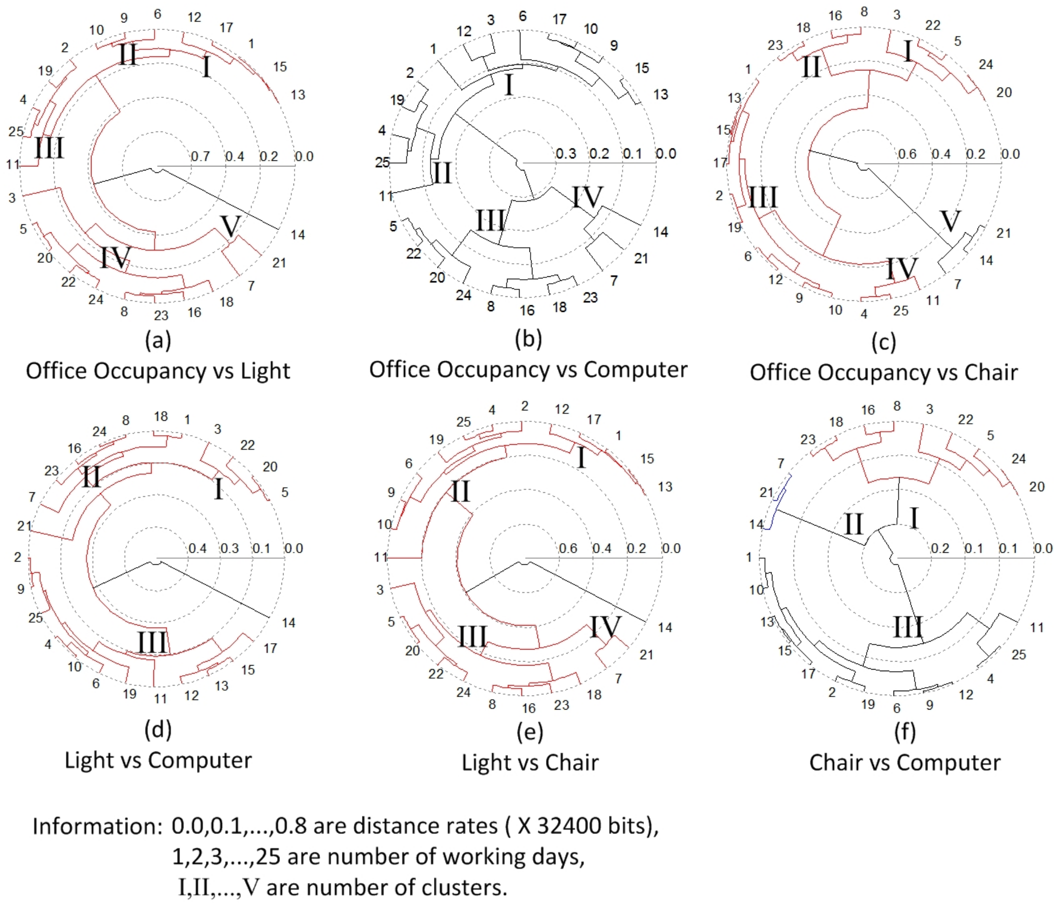



Figure 5.13: Similarity dendrogram of activities for User #2 over 25 working days.

Data set $D5$ of User #2 is used to evaluate this technique. Only twenty five working days over five weeks are considered for this investigation. Figure 5.13 shows polar-dendrograms of similarity between the activities of User #2 over twenty five working days in an office environment. In this figure, the polar-dendrogram shows similarities for six comparisons of activities for User #2. As shown in Figures 5.13-a and 5.13-c, the polar-dendrograms of similarity of office occupancy versus light activity and chair activity contained five clusters. In Figures 5.13-b and 5.13-e, office occupancy versus computer activity, and light activity versus chair activity, contained four clusters. Meanwhile, computer activity versus light activity, and light activity versus chair activity shown in Figure 5.13-d and Figure 5.13-f respectively containe three clusters.

Table 5.4 shows the similarity data clusters, translated from Figure 5.13. Clustering results are used to show a user's behaviour based on similar days. For example, it is found that for this data set office occupancy versus light activity, for days 1, 12, 13, 15, and 17 are homogeneous; i.e., the user activities (for office occupancy versus light activity) on these days are more or less similar to each other. In another example, light activity versus chair activity on day 14 does not belong to any cluster; i.e., the activities of light and chair for User #2 on day 14 are dissimilar to other days.

As mentioned earlier, the pairwise similarity data is normalised before performing the PCA Data's visualisation. This will help to illustrate the relationship between explanatory variables (user's activities) and their contribution to individual principal components [164]. For example, in Figure 5.14 the similarity score plot for the first two principal components are shown. The new scores defined by PCA are values that consist of the coordinates of the original similarity scores.

Table 5.4: Cluster of days with similar activities for User #2.

| Cluster | Office Occupancy vs Light | Office Occupancy vs Computer | Office Occupancy vs Chair |
|---|---|---|---|
| I | Day; 1,12,13,15,17 | Day;1,3,6,9,10,12,13,15,17 | Day;3,5,20,22,24 |
| II | Day;6,9,10 | Day;2,4,11,19,25 | Day;8,16,18,23 |
| III | Day;2,4,11,19,25 | Day;5,8,16,18,20,22,23,24 | Day;1,2,4,6,9,10,12,13,15,17,19 |
| IV | Day;3,5,8,16,18,20,22,24 | Day;7,14,21 | Day;4,11,25 |
| V | Day;7,21 | | Day;7,14,21 |
| **Cluster** | **Light vs Computer** | **Light vs Chair** | **Chair vs Computer** |
| I | Day;3,5,20,22 | Day;1,12,13,15,17 | Day;3,5,8,16,18,20,22,23,24 |
| II | Day;1,7,8,16,18,23,24 | Day;2,4,6,9,10,11,19,25 | Day;7,14,21 |
| III | Day;2,4,11,19,25 | Day;3,5,20,22,24 | Day;1,2,4,6,9,10,11,12,13,15,1719,25 |
| IV | | Day;7,21 | |
| V | | | |

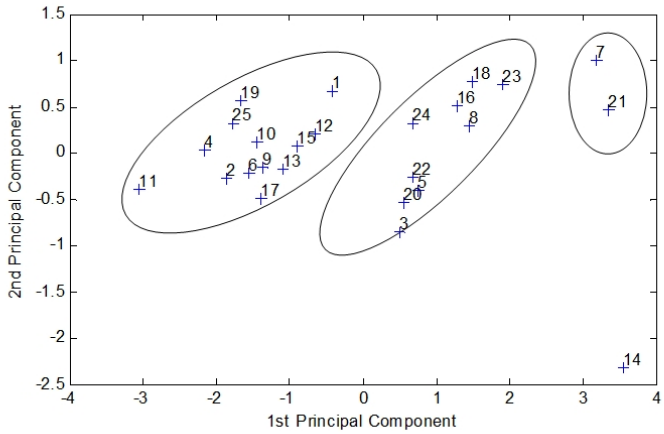
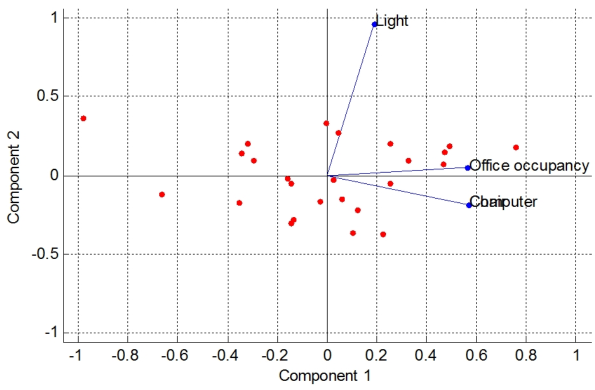


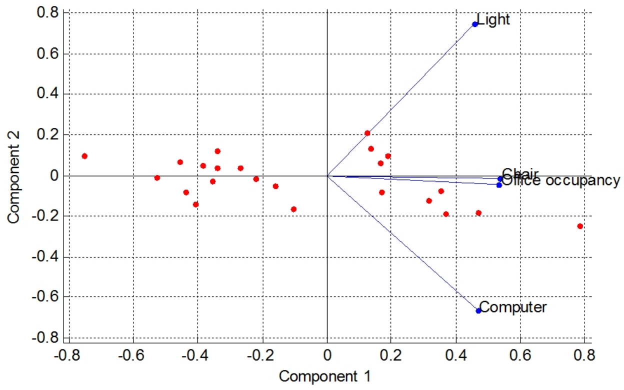

Figure 5.14: PCA's score plot for similarity between office occupancy and light activity of User #2 over 25 working days.

Figure 5.14 shows the PCA's scores between office occupancy and light activity for the $1_{st}$ and $2_{nd}$ principal components. Activities between office occupancy and lighting of User #2 on day 14 are not similar to other days. These results are the same as Dendrogram clustering results shown in Figure 5.13-a. It should be noted that only two different activities of a user can be plotted together. To be able to visualise all activities and principal components in a single plot Biplot is used. Figure 5.15 shows that each of the four activities of a user can be plotted together with the $1_{st}$ and $2_{nd}$ principal components. In order to visualise a user profile in more understandable detail, in order to contribute to an office's energy performance, the direction and length of the vector for each activity in Biplot is used to present how far each activity had contributed to the 1st and 2nd principal components [165].

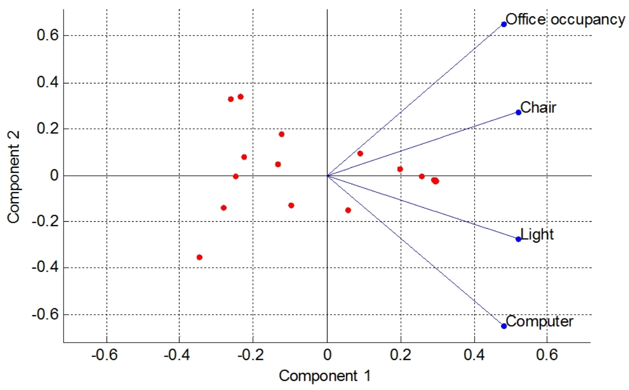
In Figure 5.15-a, the chair activity for User #1 is relatively similar to computer activity, because as the Biplot shows, both activities had the same length and direction. The office occupancy for User #1 is relatively similar to chair and




(a)




(b)




(c)

Figure 5.15: Energy use behaviour of a) User #1, b) User #2 and c) User #4, based on similarities between the activities produced using Biplot function.

computer activity, compared to light activity. From Figure 5.15-b, which shows the behaviour of User #2, we found that office computer activity is relatively similar to chair activity, compared to light and computer activities. In Figure 5.15-c, we can see that User #4 had a dissimilar behaviour to other users. In this Biplot graph, each activity of User #4 is not of the same direction or length.

**Algorithm #2: Identification of Behaviour Change**

```
function [chgs_ptime,num]=idenchgs_time(data)
w=data;
[m, n]=size(w);
j=1;chk=0;
for i=1:m-2
   if (w(i,1) ~= w(i+1,1)) && (w(i+1,1) < w(i+2,1)) && chk == 0;
      chgs_ptime(j,1) = w(i,2);
      j=j+1;
      chk=1;
   end
   if (w(i,1) ~= w(i+1,1)) && (w(i+1,1)==w(i+2,1)) && chk==1;
      chgs_ptime(j,1)=w(i,2);
      j=j+1;
      chk=0;
   end
end
chgs_ptime(:,2)= 7+(chgs_ptime(:,1)/60);
[m,n]=size(chgs_ptime)
num=m;
end
```

where: **w** is the optimal path
**chgs_ptime** is the changes positions based on time
**num** is the total number of changes

### 5.8.3 Dynamic Time Warping

Dynamic Time Warping (DTW) is used to measure similarity based on temporal warping functions and time alignment traces of sequences. The DTW curve allows us to detect changes of user activity based on time by comparing two signals. A DTW curve is formed by aligning the two sequences (temporal binary signal), where a matrix of an optimal warping path is used to plot it.

Changes in user activity are important for user behaviour identification. The proposed algorithm used to detect behaviour change based on a DTW curve is given in Algorithm #2. This algorithm is used to identify behaviour changes based on time according to a DTW curve (optimal warping curve), as shown in Figure 4.2.

Data sets $D4$ and $D5$ are used for our investigation in this section. Figure 5.16 shows the warping curve pattern and distance for office occupancy of User #1 as a reference and office occupancy of User #2 as a target. The reference sequence is in the vertical position, and the test sequence is in the horizontal position. The contour colour map of DTW illustrates the optimal path.

DTW curve is a similarity measure between motions of two sequence values on the time axis. Therefore, for those comparisons (i.e., between Day 1 and Day 3), it is found that the DTW formed the vertical and horizontal lines for inactivity and changed to diagonal lines when activity is detected. Using Algorithm #2, DTW could provide more information, such as change detection based on time for user behaviour and compute a number of changes for each comparing temporal binary signal. Arrow lines in Figure 5.17 show that the time positions where change activities between Day 1 and Day 2 for office occupancy of User #2, are

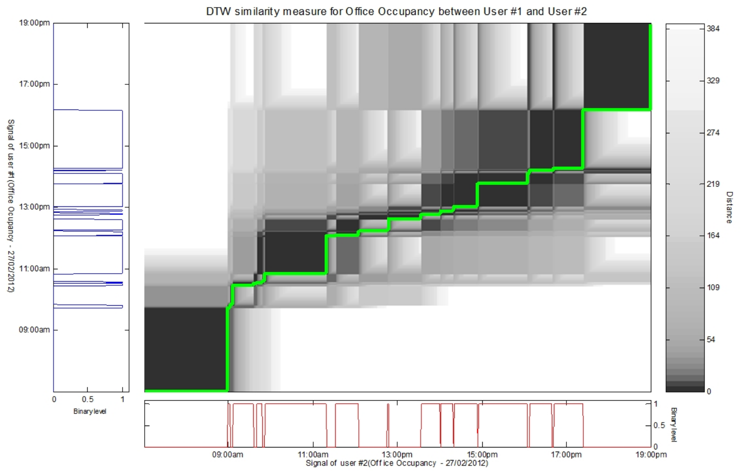



Figure 5.16: DTW similarity measure for office occupancy of User #1 as a reference and office occupancy of User #2 as target.

detected according to the DTW curve (diagonal line).

In addition, Figure 5.18 shows the summarised results of the DTW curves comparing the similarity of office occupancy for User #2 over five working days. From Figure 5.18-a, it can be seen that DTW curves can lead to dramatically different patterns, due to the comparing of signals. The DTW curves clearly show the different patterns for all combinations. In an environment control system, the DTW's capability can be used to recognise the similarity of user's behaviour and identify the real user. In Figure 5.18-b, the unnormalised distance values of DTW perform as a linear similarity measure. These unnormalised distance values are the minimum of the cumulative distance, based on the warping matrix (i.e., the calculation based on the Euclidean distance) [156]. This shows how close and far between two temporal binary signals. However, DTW allows a more intuitive distance measure to be computed than other similarity measures [167].

Moreover, identification of behaviour change is able to be detected by DTW on dynamic behaviour over a period of time. For example, Figure 5.18-c shows

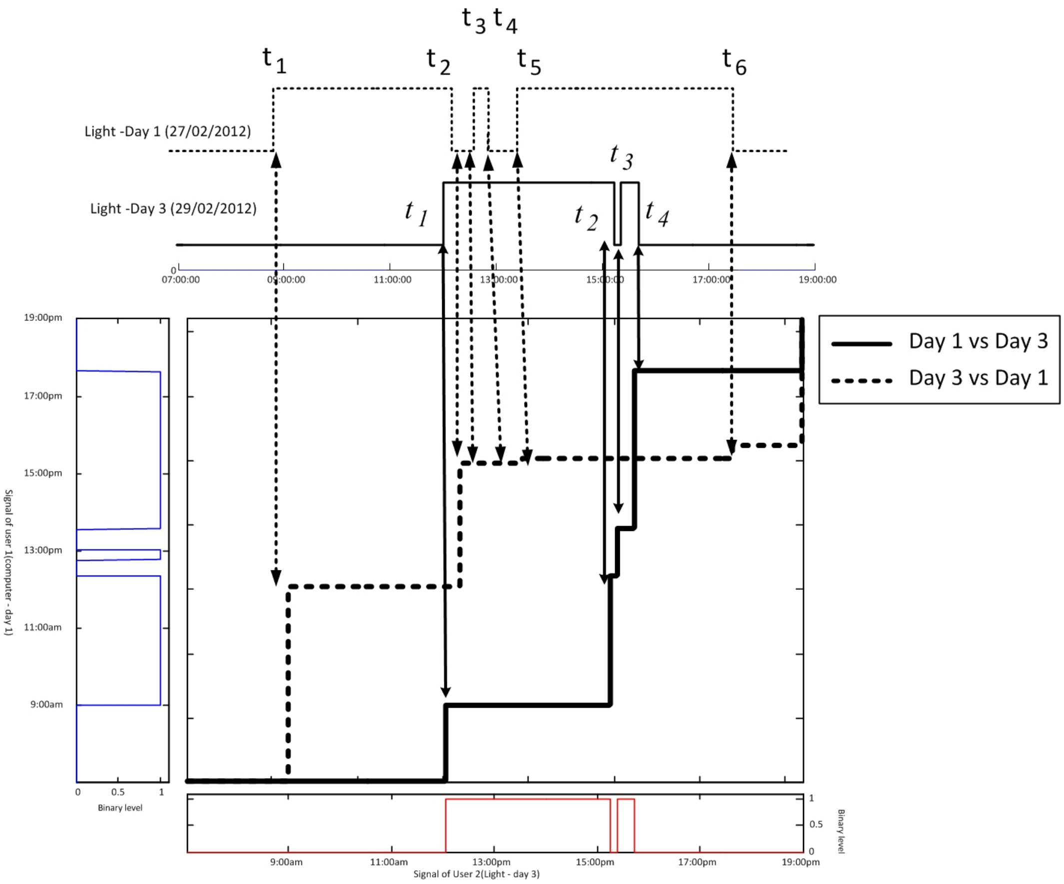



Figure 5.17: DTW curves comparing the similarities of the office occupancy between Day 1 and Day 3 of User #2.

the results of how Algorithm #2 detected the positions of time, based on behaviour changes over a period of time. This contributes to improvements in the sensitivity of an environmental control system to users' dynamic behaviour. Furthermore, Figure 5.18-d also shows the capability of DTW to detect the number of changes/positions based on time. Therefore, it would also contribute to increases in the intelligence of an environmental control system to recognise real users or change the control mode suitable to the current behaviour of the user (e.g. based on chaotic levels and consistent levels as discussed in Chapter 4).

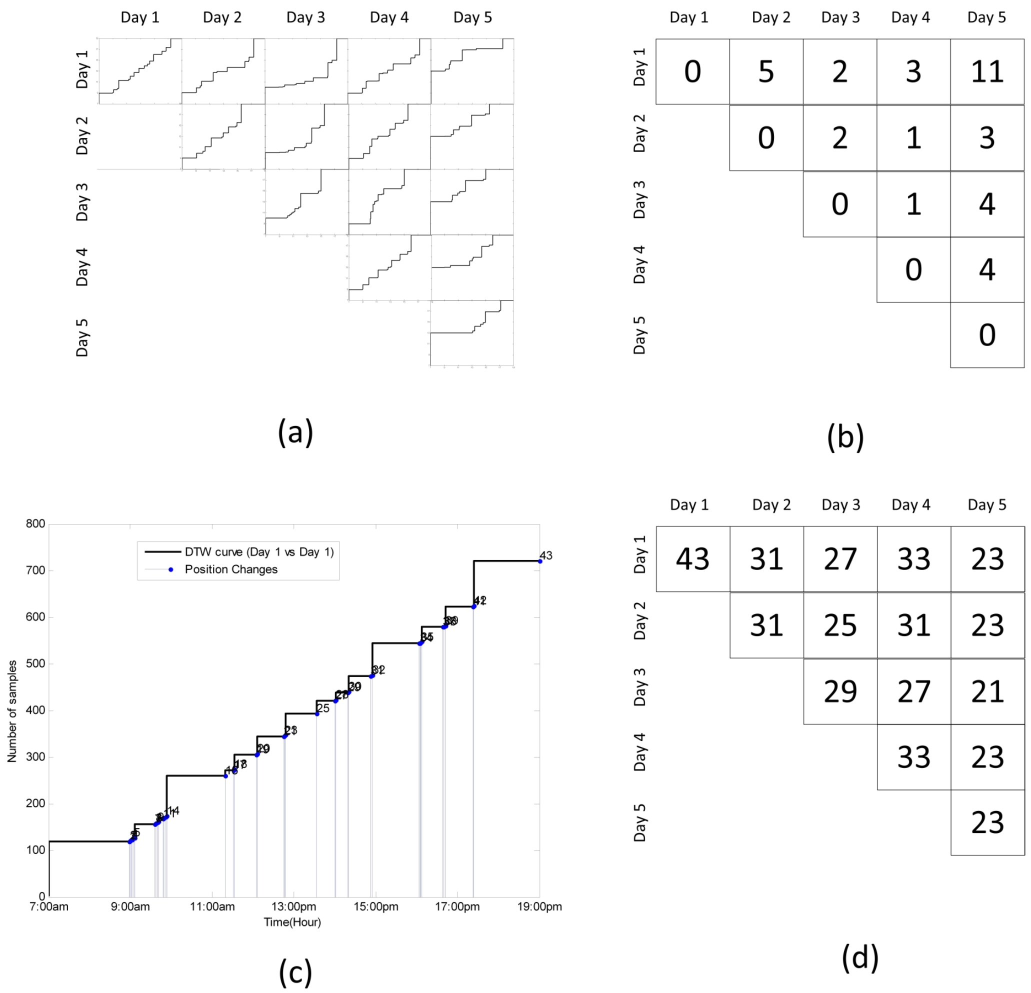



Figure 5.18: DTW results for user #2 over five working days. a) DTW curves comparing the similarity of office occupancy b) distances between binary signals for office occupancy c) office occupancy behaviour changes over time d) number of changes/positions.

## 5.9 Discussion

This chapter presents the experimental results used to analyse and extract user profiles from the user's behaviour data. Time-series signals of user activities and ambient conditions in an office environment are recorded using a data collection system. However, these signals provided little information to detect behaviour;

and thus construct a user profile. Several experiments in this chapter are successfully implemented, and methods used to analyse data from office worker's behaviour in an office environment are proposed. In order to deal with large ADW binary signals, several progressive methods of data representation and converting sensory signals into multidimensional knowledge are proposed. Converting data into start-time and duration sequences, and signal unification, contributed to discovering user behaviour as follows:

- By reducing the dimensionality of the stored data, it is able to help convert the long period ADW data into a more understandable format.
- Minimising the distances between the sequence data of sensory signals in bit order, in order to generate a high accuracy of prediction, would offer the unification signal as an enhancement tool for the prediction model, in order to speed up and reduce errors.

The patterns of users' behaviour and office's thermal comfort in a statistical perspective are used to extract information to construct user's profiles. In this chapter, these patterns are evaluated using comparative methods using numerical statistical values and graphs. For example, Cronbach's ($\alpha$) is used to measure internal consistency and ApEn is used to measure the chaos of user's behaviour. Tabular and graphical statistical representations are also used to extract individual user profiles, as proposed and shown in Figure 5.1.

Statistical analysis experiments in this chapter detected that each office user had different behaviours. The ADW of office workers had different patterns on different days of the week. For example, the results shown in Table 5.5 indicate that User #1, #2, #4 are dissimilar habitually at the start of work, where User

#1 is the earliest to arrive at the office, and User #2 left the office for the day later than other users. In terms of user activity duration, due to different schedules of work, User #2 spent more time in the office with above average computer usage than User #1 and #4. Based on the consistency and chaotic measures of user's activity, we found that the consistency of computer activity for User #1 is inconsistent with the other users, with an $\alpha$ of less than 0.5, and the chaotic levels of all activities of user's is moderate, with an ApEn of less than 0.8. According to the results, each office user's working attitude and preferences could be summarised in the profile properties (as shown in Table 5.5). Table 5.6 shows the formulas used to extract users' profile information from the users' behaviour data.

Table 5.5: User profiles, used for summarizing user's behaviour in an office environment over twelve weeks.

| User | Working Time | | Avg. Duration/Routine habit of activity in an office | | | | Thermal Comfort |
|---|---|---|---|---|---|---|---|
| | Arrival | Departure | Occupancy | Lighting | Chair | Computer | Temperature |
| #1 | 08:33 am | 17:43 pm | *M/(C,P)* | *M/(C,P)* | *M/(C,P)* | *M/(InC,P)* | 19 to 23 |
| #2 | 09:13 am | 19:13 pm | *L/(C,P)* | *L/(C,P)* | *M/(C,P)* | *L/(C,P)* | 22 to 25 |
| #4 | 09:11 am | 18:32 pm | *M/(C,P)* | *M/(C,P)* | *M/(C,P)* | *S/(C,P)* | 20 to 25 |

Information: 1. The levels of duration are indicated such as $L$ (long), $M$ (medium) and $S$ (short).
2.The levels of measures for consistency and choatic are indicated as follows:
i. Consistency - $C$-consistent and $InC$-inconsistency
ii. Chaotic - $P$-predictable/less chaotic and UnP-unpredictable/chaotic

Table 5.6: Formulas used to extract profile information from the user's behaviour data.

| Working Time | Duration /Routine habit of activity in an office | | | Thermal Comfort |
|---|---|---|---|---|
| | Avg. Duration | Consistency | Chaotic | Temperature |
| Arr.=$min(t_{door})$ | $L > 25\frac{hours}{week}$ | $\alpha$ | ApEn | $min(tmp)$ to $max(tmp)$ |
| Dept.=$max(t_{door})$ | $15\frac{hours}{week} < M < 25\frac{hours}{week}$ | $C$:$\alpha > 0.5$ | $P$: $ApEn < 0.6$ | $tmp$=Tmp. for moving |
| | $0\frac{hours}{week} < S < 15\frac{hours}{week}$ | $InC$:$\alpha < 0.5$ | $UnP$: $ApEn > 0.6$ | avg. of 5 data points |

According to the experimental results of measuring similarities of behavioural patterns shown in Table 5.3, Faith and HD demonstrated that they worked better for all binary signals at measuring the similarity of activities. The other methods showed less sensitivity to slight deviations for similarity degrees between two different activity sequences. HD is selected to measure the similarity of users' behaviour for a period of time. The results shown in Figure 5.16 showed that each user had a different behaviour, while working in an office environment. Therefore, the pairwise presentation for similarity between different user's activities is proposed. By applying this method, similarity scores are used to cluster the user's activities using a (polar) Dendrogram plot, where the data is structured using hierarchical clustering. Table 5.4 shows which user activities are relatively similar to other activities on different days; based on clustering results. Data visualisation, using PCA and Biplot, is proposed. Figures 5.14 and 5.15 are used to help illustrate the relationship between different user' activities and their contribution to principal components.

The DTW, as a non-linear similarity measure, is used to measure similarity and identify dynamic behaviour change over time. Different user's activities are compared using DTW , where the optimal match curve indicated the degree of similarity between two different user's activities in the time dimension (see Figure 5.16). The experimental results (see Figure 5.17 and 5.18) show that DTW is more applicable to compare the similarity between two users than the linear similarity measures. An optimal match curve is an adjoining set of matrix elements that is described as a mapping between two different sequences of activity [167]. Algorithm #2 is proposed to detect behaviour change over time. Through a combination of DTW's algorithm and Algorithm #2, a DTW curve can provide more information, such as similarity based on the optimal warping curve; distance value, identification of behaviour changes, and a number of computed changes/positions based on time (see Figure 5.18).

In conclusion, the individual user's profile (as shown in Table 5.5) can be used to distinguish between different user's profiles. Moreover, repetitive regular behaviour patterns can be detected using the similarity algorithms. This would benefit an office environment control system, by enabling it to recognise real users and update the user's profile. This would also enable the office environment to be automatically adjusted in line with the user's preferences. The control system could be improved; if it is able to identify the similarities and changes of user's behaviour over a period of time.

Chapter 6 will describe the techniques of enhanced user profiling and ADW recognition in an intelligent office environment.

# Chapter 6

# Enhanced User Profiling

## 6.1 Introduction

To have an effective and efficient control strategy in an office environment, it is required to understand the uncertainties involved in a user behaviour. In the previous chapter, an attempt was made to use the statistical measures as a representative of the behaviour. There are alternative approached to the statistical techniques and soft computing techniques have proven to be a useful to tackle the challenges of this area of research. Specifically, fuzzy logic and fuzzy rule-based systems are a proper tool to model uncertainties.

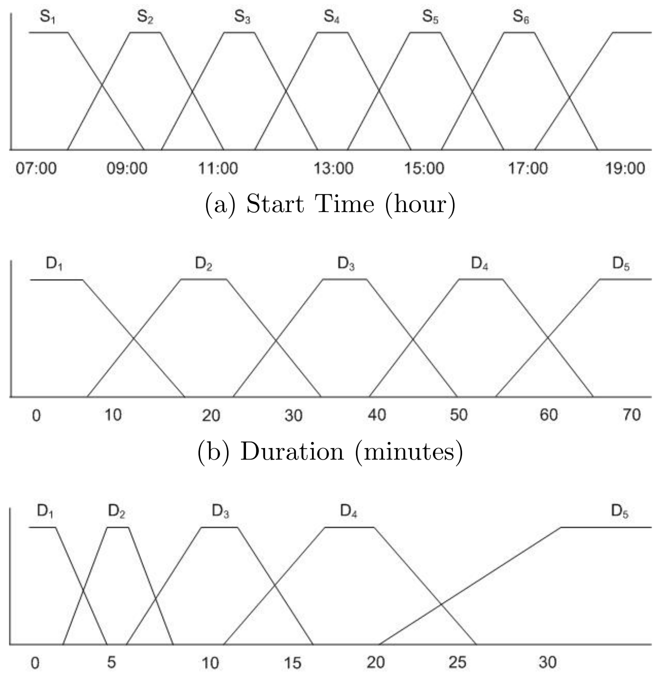
The aim of this chapter is to incorporate soft computing techniques namely fuzzy system to identify ways to enhance the user profiling mentioned in the previous chapter. To achieve this goal, it is important to understand each user's profile initially before adapting the environment to the users requirements. For example, instead of waiting half an hour for a computer auto shutdown, the system could detect that the user is no longer in the room and the computer is not performing any work. This would cause the office environment control to put the PC into an energy saving mode.

The rest of the chapter is structured as follows: an overview of the fuzzy characteristics matrix is presented in Section 6.2 followed by technique for generating user's fuzzy profiles in Section 6.3. In Section 6.4, an office worker's ADW recognition using event-driven model is presented. In Section 6.5, an ADW recognition based on fuzzy rule-based system is explained and the experimental results are also discussed. Finally, conclusions are drawn in Section 6.6.

## 6.2 Fuzzy Characteristics Matrix

Collected data from each office represent the user profile for that environment. Different users have different preferences and it is essential to summarise the information into meaningful knowledge within the control server application to control each office environment effectively. To process the collected data, two stages are identified. Firstly, signals are presented in start-time, $ST_k(t)$, and duration, $DU_k(t)$, sequences as discussed in Section . Converting the raw data into $ST_k(t)$ and $DU_k(t)$ for each activity, $k$, will summarise the data and present them in a compact format. In the second stage, these sequences are fuzzified.

Knowing the exact time (in seconds or even minutes) for the start-time and duration of an activity would not be very significant. For example if an office worker is usually come to his/her office early morning, then this information should be enough to create a user profile without specifying the exact time every day. Therefore, in order to manipulate categorical data rather than numerical data, $ST_k(t)$ and $DU_k(t)$ for all sensor readings are replaced by their membership




(a) Start Time (hour)



(b) Duration (minutes)



(c) Duration (minutes)

Figure 6.1: Fuzzy partitions a) start time for all events b) duration for chair occupancy c) duration for PC activities.

in several fuzzy sets.

Time of the day is split into $m = 7$ overlapping fuzzy sets, $\tilde{S}_i, i = 1, ..., m$. They are $\tilde{S}_1$: EarlyMorning, $\tilde{S}_2$: Morning, $\tilde{S}_3$: LateMorning, $\tilde{S}_4$: EarlyAfternoon, $\tilde{S}_5$: Afternoon, $\tilde{S}_6$: LateAfternoon, $\tilde{S}_7$: Evening. Fuzzy partition for $ST_k(t)$ is shown in Figure 6.1-a.

Duration of events is fuzzified on an event basis. Working in front of a PC for a "short time" is not the same as opening the window for a "short time". Figure 6.1-b and Figure 6.1-c illustrate fuzzy values for duration of chair occupancy and PC activities respectively. For all events, duration is split into $n = 5$ overlapping fuzzy sets, $\tilde{D}_j, j = 1, ..., n$. They are $\tilde{D}_1$: VeryShort, $\tilde{D}_2$: Short, $\tilde{D}_3$: Medium, $\tilde{D}_4$: Long, $\tilde{D}_5$: VeryLong.

The degree of membership, $\mu$ (which is the degree to which the event satisfies the $ST_k(t)$ and $DU_k(t)$ fuzzy labels) are calculated and these grades will be used to create a profile for a user.

## 6.3 User's Fuzzy Characteristics

Start time and duration of each event are represented in membership values for all fuzzy labels. Therefore for each event, two arrays of membership values for $ST_k(t)$ and $DU_k(t)$ are produced.

$$[\mu_{\tilde{S}_i}]\, and\, [\mu_{\tilde{D}_j}],\, for\, i = 1, \ldots m; j = 1, \ldots n \tag{6.1}$$

Fuzzy values collected for a sensor are combined to form a matrix of membership values. For example all events related to the occupancy sensor are fuzzified and two matrices of fuzzy values for start time and duration are generated. This is summarised in the following expression:

$$\Sigma_k = \begin{bmatrix} \mu_{\tilde{S}_1^1} & \cdots & \mu_{\tilde{S}_i^1} & \cdots & \mu_{\tilde{S}_m^1} \\ \mu_{\tilde{S}_1^2} & \cdots & \mu_{\tilde{S}_i^2} & \cdots & \mu_{\tilde{S}_m^2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \mu_{\tilde{S}_1^p} & \cdots & \mu_{\tilde{S}_i^p} & \cdots & \mu_{\tilde{S}_m^p} \end{bmatrix}_k \tag{6.2}$$

$$\Delta_k = \begin{bmatrix} \mu_{\tilde{D}_1^1} & \cdots & \mu_{\tilde{D}_j^1} & \cdots & \mu_{\tilde{D}_n^1} \\ \mu_{\tilde{D}_1^2} & \cdots & \mu_{\tilde{D}_j^2} & \cdots & \mu_{\tilde{D}_n^2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \mu_{\tilde{D}_1^p} & \cdots & \mu_{\tilde{D}_j^p} & \cdots & \mu_{\tilde{D}_n^p} \end{bmatrix}_k \tag{6.3}$$

where $p$ the total number of occurrences for a specific sensor, $k$. Each row of the matrix $\Sigma_{p\times m}$ and $\Delta_{p\times n}$ are membership values for start time and duration of and event respectively. Number of rows very much depends on the activity and the user profile.

To create a user profile based on each activity, it is recommend to create a fuzzy characteristics matrix, $\Phi$:

$$\Phi_k = \Sigma_k^T \times \Delta_k \tag{6.4}$$

where $\Phi_k$ is a $n \times m$ matrix and this can be generated for all events and activities.

The observation of collected data shows that worker's behaviour patterns are not fixed and may vary on different days of the week; depending on the nature of the work. Figure 6.2 shows a sample of activity's duration over three weeks for User #2. The solid lines show the daily moving average for the same dataset. The worker clearly has different patterns on different days of the week. Therefore, the coarse profile only considers the average in order to provide the first stage of modelling. There will be some seasonal changes; therefore, that pattern of work may vary during summer and winter.

To illustrate the previously discussed concept of a fuzzy characteristics matrix, the room occupancy of a user was considered for two separate days. The occupancy sensor was activated and deactivated many times during these working days. Fuzzy characteristic matrices are created for both days, using the data shown in Figure 6.3. In order to visualise these profile matrices, they are displayed using scatter plots. Figure 6.4 compares the worker's behaviour; emphasising dominant patterns and clearly differentiating between different patterns

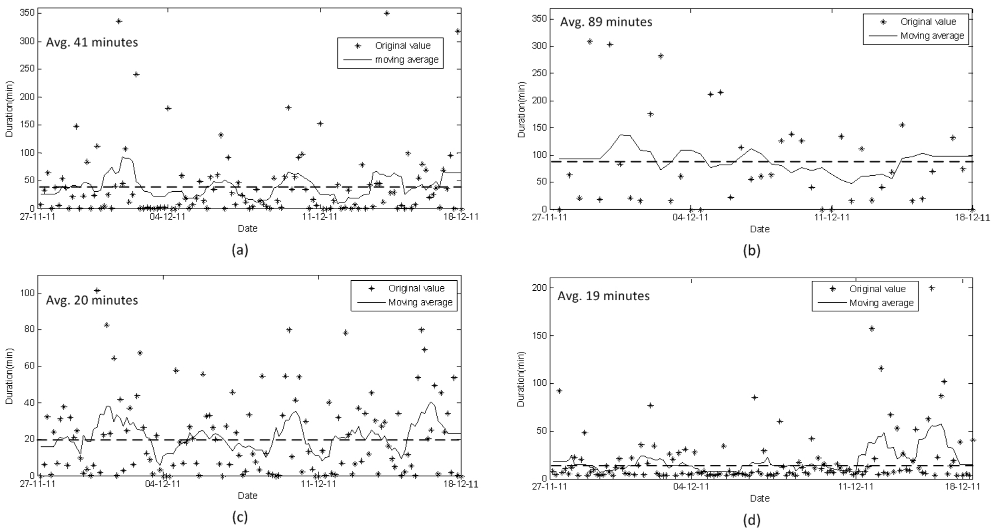



Figure 6.2: Sample of activities duration of User #2 over three weeks for a) office occupancy, b) lighting, c) chair occupancy, and d) computer usage.

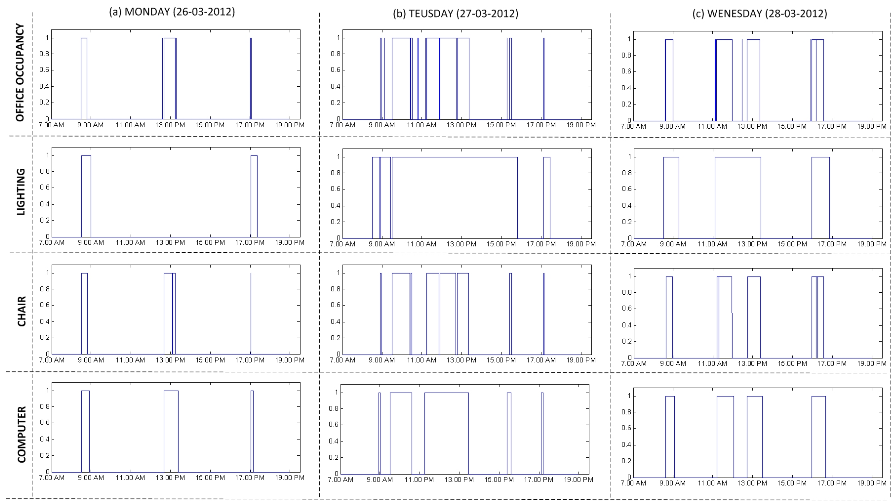



Figure 6.3: Sensor activities for User #2 on a) Monday, b) Tuesday, and c) Wednesday.

of use.

In these plots, the size of the circle represents the likelihood of that activity occurring at the specified start time and the duration for that day of the week. Different colours are used to represent the different activities of the user. For

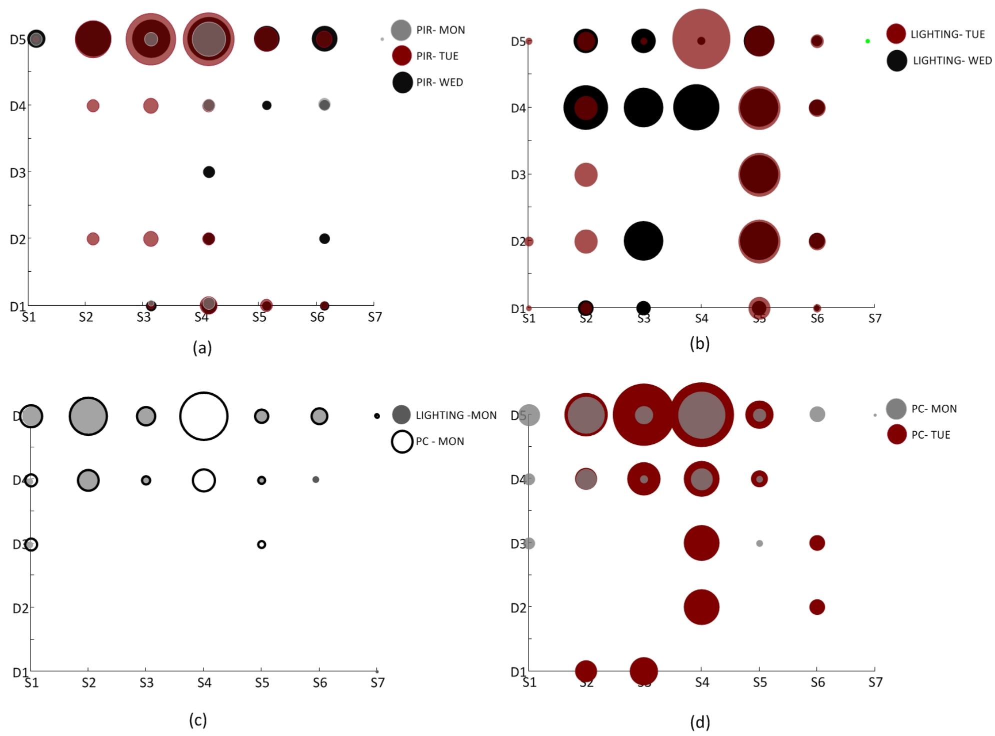



Figure 6.4: Scatter plot for fuzzy characteristics matrix a) Office occupancy for Monday, Tuesday, and Wednesday, b) Lighting for Tuesday and Wednesday, c) between Lighting and Computer, and d) Computer for Monday and Tuesday.

example, the results obtained from the comparison between lighting and computer usage on Monday are shown in Figure 6.4-c. In this figure, the user's activities for light usage and computer usage have similar durations at $S_1$: EarlyMorning, $S_2$: Morning, $S_3$: LateMorning, $S_5$: Afternoon, and $S_6$: LateAfternoon. However, different durations appear at $S_4$: EarlyAfternoon, where the user used a computer for a long duration, without using light usage on Monday.

In the next section, activity recognition will be discussed to identify the activity of an office worker; whether the user's activities are in the office or out of the office, during a typical working day.

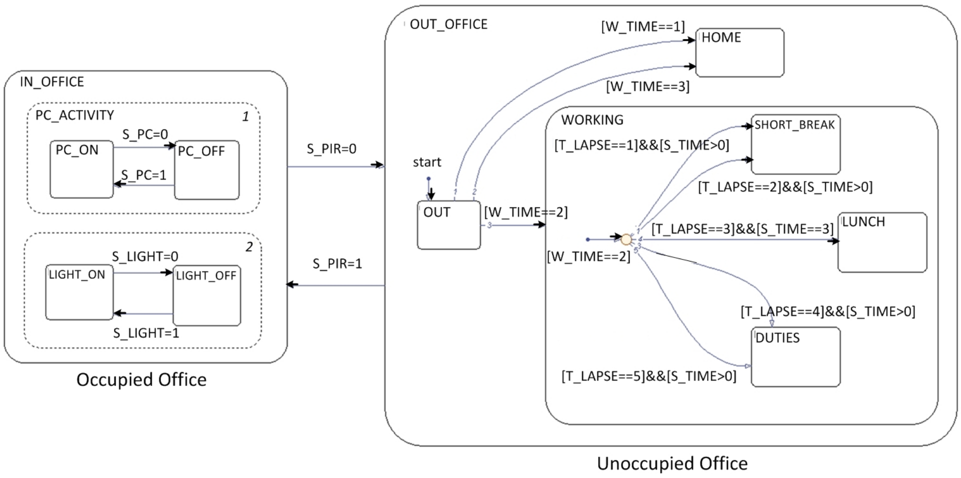



Figure 6.5: The events flow hierarchy of an event-driven model to discover user's activities in an office environment.

## 6.4 Activity Recognition

In order to identify ADW of an office worker, it is important to distinguish between the time the office is occupied and unoccupied. It is also important to distinguish if the office is unoccupied due to breaks/out-of-office duties, and the time that the office is unoccupied due to leaving at the end of the work day. This can include the time that the user spends on out of the office to perform duties such as teaching, meetings, etc. An event driven approach is applied to guide transitions, where events are used to indicate a certain scenario of a worker's activity.

### 6.4.1 An Event-Driven Approach

The events flow of an office worker is presented in a hierarchy diagram. Figure 6.5 shows the hierarchy of events flow for an individual office worker. In order to define the event of user's activity, the input, output and parent of an event are specified. Secondly, the events flow is determined based on a worker's ADW in the office and out of the office. The main event is separated into two separate events of IN OFFICE and OUT OF OFFICE.

Based on the information gathered from the occupancy sensor (PIR) , the event will be determined. These two events hold many other events based on the conditions of other sensors.

Sensor conditions are used to control the event sequence in a transition within an event driven model. In addition to sensor conditions, three temporal variables are also defined. They are; *Time Lapse*, *Start Time* and *Working Time*. Time Lapse is defined as an interval of time where a worker is out of the office. It starts when the worker leaves the office and ends when the worker returns. A variable of Start Time is used to indicate when the worker leaves the office and returns based on time partitions (e.g. morning, afternoon, evening). The variable of Working Time is divided into three categories i.e. before start work, working and finish work. It indicates the period of time that an individual worker arrives

Table 6.1: Temporal variable's values and ranges for activities recognition model.

| Temporal Variable | Value/Range |
|---|---|
| Time Lapse | $T\_LAPSE = \{\ Very\_Short$=1, $Short$=2, $Medium$=3, $Long$=4, $Very\_Long$=5 |
| Start Time | $S\_TIME = \begin{cases} \text{Early Morning}(EM); & \text{IF 7.00 AM} < EM < \text{8.30 AM} \\ \text{Morning}(M); & \text{IF 8.30 AM} < M < \text{10.30 AM} \\ \text{Late Morning}(LM); & \text{IF 10.30 AM} < LM < \text{12.00 PM} \\ \text{Early Afternoon}(EA); & \text{IF 12.00 PM} < EA < \text{13.00 PM} \\ \text{Afternoon}(A); & \text{IF 13.00 PM} < A < \text{15.00 PM} \\ \text{Late Afternoon}(LA); & \text{IF 15.00 PM} < LA < \text{17.00 PM} \\ \text{Evening}(E); & \text{IF 17.00 PM} < E < \text{19.00 PM} \end{cases}$ |
| Working Time | $W\_TIME = \begin{cases} \text{Before Arrival= 1;} & \text{IF } TimePIR_i\ <\ start\ of\ W\_TIME \\ \text{Working= 2;} & \text{IF } (TimePIR_i\ >=\ start\ of\ W\_TIME) \\ & AND\ (TimePIR_i\ <=\ end\ of\ W\_TIME) \\ \text{After Departure= 3;} & \text{IF } TimePIR_i\ >\ end\ of\ W\_TIME \end{cases}$ |
| | where, $TimePIR_i$ is the time of PIR's sensor is activated, and $i$=1,2,...,n. |

and departs from the office.

Since worker activity and the real time are recorded by an intelligent office environment system, the granularity of temporal variables are generated using statistical algorithms based on office occupancy signals from the PIR sensor. A worker's ADW patterns are not fixed and they may vary on different days of a week depending on the nature of the work. The values of time lapse are constructed using the semi-interquartile range [78]. Based on previous work in Section 6.3, start time is divided into 7 partitions, working time is divided into 3 partitions and time lapse is divided into 5 partitions. A working time is defined as the interval between the arrival and departure times of the worker to the office. These times are detected by office occupancy recorded from a PIR sensor. The statistical algorithms provided the following data outputs for the temporal variables values as shown in Table 6.1.

A set of heuristic If-Then rules is applied to control the transition of events for event-driven activity recognition. An algorithm consists of the rules to determine the event's transition of worker's activities based on sensory signals. The sensory signals and temporal variables are used as the antecedence part of the rules, where the consequence part are expressed as a sequence pattern of $X$ as shown below with indicated possible values.

$$X = \{HOME, DUTIES, SHORT\_BREAK, LUNCH, PC\_OFF, PC\_ON\}$$

Linguistic rules are listed below:

1. **IF** $S\_PIR_i = 0$ **AND** ($W\_TIME = 1$ **OR** $W\_TIME = 3$) **THEN** $X = HOME$.

2. **IF** $S\_PIR_i = 0$ **AND** $W\_TIME = 2$ **AND** $S\_TIME > 0$ **AND** (($T\_LAPSE = 4$ **OR** $T\_LAPSE = 5$) **THEN** $X = DUTIES$.

3. **IF** $S\_PIR_i = 0$ **AND** $W\_TIME = 2$ **AND** $T\_LAPSE = 3$ **AND** $S\_TIME = 3$ **THEN** $X = LUNCH$

4. **IF** $S\_PIR_i = 0$ **AND** $W\_TIME = 2$ **AND** $S\_TIME > 0$ **AND** (($T\_LAPSE = 1$ **OR** $T\_LAPSE = 2$) **THEN** $X = SHORT\_BREAK$.

5. **IF** $S\_PC_i = 1$ **THEN** $X = PC\_ON$.

6. **IF** $S\_PC_i = 0$ **THEN** $X = PC\_OFF$.

7. **IF** $S\_Light_i = 1$ **THEN** $X = LIGHT\_ON$.

8. **IF** $S\_Light_i = 0$ **THEN** $X = LIGHT\_OFF$.

where $S\_PIR_i$ represents office occupancy from PIR sensor, $S\_Light_i$ represents light sensor and $S\_PC_i$ represents computer sensor. Subscript $i$ denoted the time in the sequence of data set. A temporal variable is denoted such as $W\_TIME$ for working time, $S\_TIME$ for start time and $T\_LAPSE$ for time lapse.

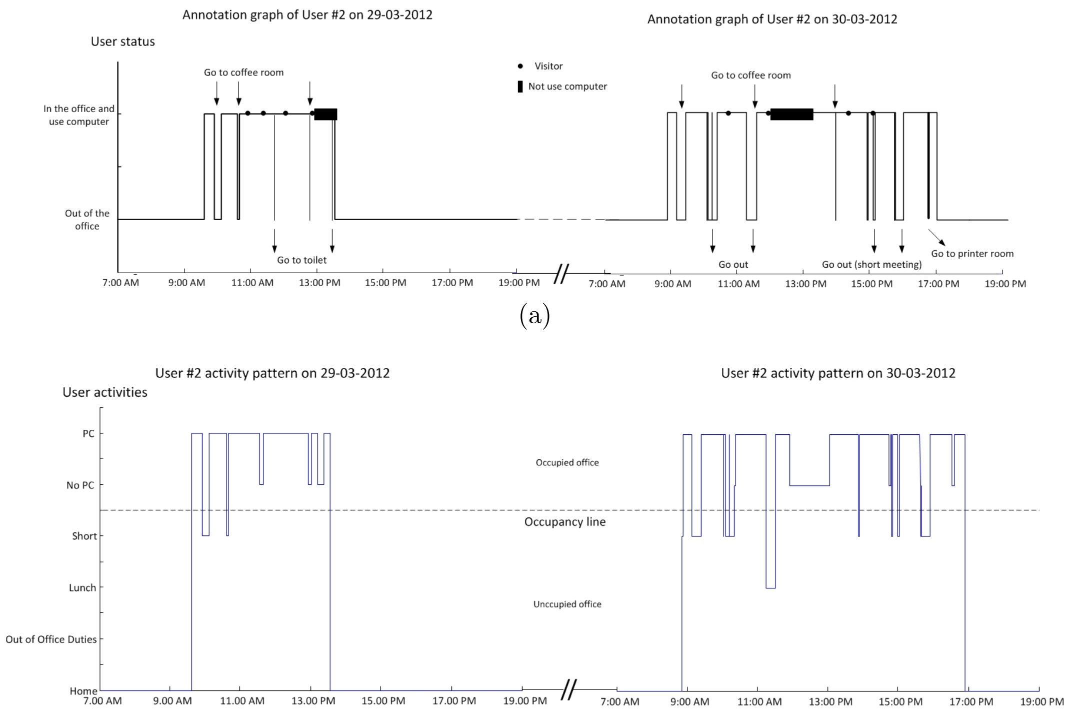
A time lapse variable is used to determine events when an office is not occupied. This is represented as $DUTIES$ for out of office duties, $SHORT\_BREAK$ for a short break and $LUNCH$ for lunch period. For example, if $S\_PIR = 0$, then OUT state is activated. The destination of transition is determined by $W\_TIME = 1$, $W\_TIME = 2$ and $W\_TIME = 3$. If $W\_TIME = 2$ is true, the transition to WORKING state is valid. In this state, the destinations of transitions are determined by $T\_LAPSE$. If $T\_LAPSE = 2$ and $S\_TIME = 3$ is true, the transition is to LUNCH. The LUNCH state will be active until $S\_PIR = 1$, where a worker has entered back to the office. Temporal variables are used to determine whether the worker has gone home, having a short break, on lunch break or is performing out of duties.

### 6.4.2 Implementation and Results

To test and validate our approach to recognise an office worker's activities, collected data from our experimental offices are initially annotated [168, 169]. A sample of user annotated records and validation results are shown in Appendix F. Annotation of the collected data will help to understand the relationship between the raw data and the actual activity. The rules generated from the sensory data are compared against the actual activity.

Sensor based activity recognition using an event-driven and If-Then rules approach have demonstrated success in tracking the activities of a worker during working days. Data set $D5$ of User #2 is used. Based on annotation of ADW from User #2, the activity pattern is shown in Figure 6.6-a. Figure 6.6-b represents the output activity patterns of the same user recognised by developed model. The dashed line is used to divide occupancy into two parts: the occupied office part and unoccupied office part. The worker clearly has different patterns on different days. The proposed activity recognition system is able to discover patterns for a worker's activities such as short break, lunch and out of office duties.

Output of this model are used to describe a user's profile and pattern of




(a)

(b)

Figure 6.6: Samples of activities patterns based on a) user's annotation b) user's activity model gathered from sensory devices.

computer usage. Comparing the PC activities with the recognition signal, it is possible to identify periods when the PC could have been switched off. With an accuracy of 91%, the recognition model has identified the PC activities. However, the accuracy will be even better if the recognition signal is compared with PIR signal. This could be up to 99% accuracy. More details of the validation of activity recognition for users' behaviour are presented in Appendix F.

Hence the proposed methods recognise a user's activity and computer usage pattern from sensory data such as PIR and PC. Figure 6.7-a shows the patterns comparing between occupancy, computer usage and a user's activity. The model of activity recognition are used to discover a worker's activities and describe a pattern of lighting or radiator usage by replacing the PC's sensor with light sensor

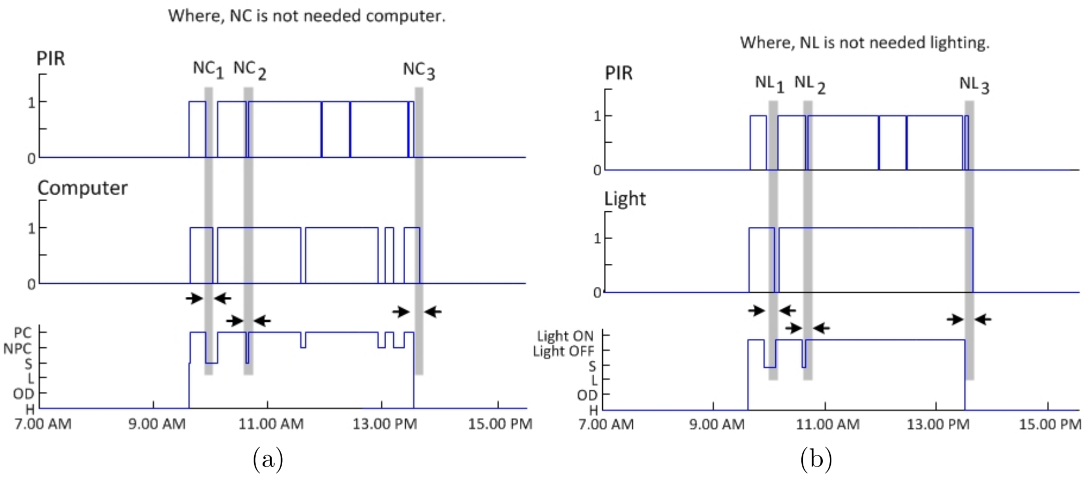



Figure 6.7: Samples of patterns for a) user's activity, office occupancy and computer usage b) user's activity, office occupancy and lighting usage.

or heater temperature sensor as the input of the model. Figure 6.7-b compares the patterns between occupancy, lighting usage and a user's activity.

Further investigations are conducted to identify a worker's activity and pattern of power usage. Detailed investigation of results in Figure 6.7 found that, after a user leaves the office, the computer and lighting are still powered on for around five to fifteen minutes. In Figure 6.7, the trace lines show the unnecessary operation of computer and lighting; $NC_1$, $NC_2$ and $NC_3$ are unnecessary operations of computer, while $NL_1$, $NL_2$ and $NL_3$ are unnecessary operations of

Table 6.2: The estimation of a user's power consumption of computer and lighting in an office for a day.

| | Computer Power Usage | | | | Lighting Power Usage | | | |
|---|---|---|---|---|---|---|---|---|
| | Total/day | $NC_1$ | $NC_2$ | $NC_3$ | Total/day | $NL_1$ | $NL_2$ | $NL_3$ |
| Use office(hour) | 3.57 | 0.1222 | 0.0611 | 0.1167 | 3.97 | 0.1833 | 0.0611 | 0.1056 |
| Power used(KWh) | 1.178 | 0.0403 | 0.0202 | 0.0385 | 0.3811 | 0.0176 | 0.0590 | 0.0101 |
| Unnecessary | 0.1 KWh | | | | 0.08 KWh | | | |
| Information: | Computer = 300 Watt<br>Monitor=30 Watt<br>Lighting= 24 Watt × 4 units | | | | | | | |

lighting detected when a user leaves the office. Table 6.2 provides the results of power consumption measurements based on the results presented in Figure 6.7. The results of the estimation measurements on computer power usage for a day is 1.178 KWh, and the power consumption of computer that is identified as unnecessary reached 0.1 KWh. The lighting power usage for a day is 0.3811 KWh, and the power consumption of lighting that is identified as unnecessary reached 0.08 KWh.

To take into account the uncertainty involved in detection of partitions, in the next section a fuzzy If-Then rule approach will be investigated.

## 6.5 Fuzzy Inference System Approach in Activity Recognition

Fuzzy Inference System (FIS) is used to give flexibility to the definition of time in quantifying the activity of a worker. The development of worker activity recognition using FIS requires the two important elements of a justifying fuzzy variable design and fuzzy rules based on the observation of a worker's activity data.

The following fuzzy variables are identified:

- **Sensors:** The sensor's values indicate office occupancy, chair sensor, and PC monitoring software, and are presented in 'ON' and 'OFF' stages, continuously over a period of time.

- **Start time:** Time of the day is split into $m = 7$ overlapping fuzzy values.

- **Time-lapse:** Time-lapse are durations where a user takes time to perform activities out of the office during a working hour. Time-lapse is the duration when the PIR sensor is inactive. The fuzzy membership function for the time lapse is constructed based on a calculation, using semi-interquartile range according to distribution data of an inactive occupancy sensor and has five symbols. Given the $25^{th}$ percentile $q_1$, median $q_2$, $75^{th}$ percentile $q_3$ and interquartile range $IRQ$=($q_3-q_1$) of data, each $\mu$F is calculated.

- **Working time:** Working time is a measure of the length of time between when the worker arrives at and departures from the office.

Using the data for two sensors (e.g. PIR and PC) combined with three temporal variables, the rules can be simplified as follows:

**Rule** $k_1$: **IF** PIR is *Inactive* **AND** PC is *Inactive* **AND** Working_Time is *Before start* **AND** Start_Time is Early morning, **THEN** *activity is Home.*

**Rule** $k_2$: **IF** PIR is *Active* **AND** PC is *Inactive* **AND** Working_Time is *Working* **AND** Start_Time is Morning, **THEN** *activity is Not use PC.*

**Rule** $k_3$: **IF** PIR is *Inactive* **AND** PC is *Active* **AND** Working_Time is *Working* **AND** Start_Time is Morning, **THEN** *activity is Use PC.*

**Rule** $k_4$: **IF** PIR is *Inactive* **AND** PC is *Inactive* **AND** Working_Time is *Before start* **AND** Time_Lapse is *Short* AND Start_Time is *Morning*, **THEN** activity is *Short break.*

**Rule** $k_4$: **IF** PIR is *Inactive* **AND** PC is *Inactive* **AND** Working_Time is *Before start* **AND** Time_Lapse is *Long* **AND** Start_Time is Morning, **THEN** *activity is Out of office duties.*

$\vdots \qquad \vdots \qquad \vdots$

**Rule** $k_i$:**IF** PIR is *Inactive* **AND** PC is *Inactive* AND Working_Time is *Before start* AND Time_Lapse is Medium AND Start_Time is *Early Afternoon*, **THEN** *activity is Lunch.*

$\vdots \qquad\qquad \vdots \qquad\qquad\qquad\qquad \vdots$

**Rule** $k_n$: **IF** PIR is Inactive AND PC is Inactive AND Working_Time is Left office **AND**, **THEN** activity is *Out of office duties.*

Possibility distribution inference [170] is applied to approximate the fuzzy activity pattern. The possibility distribution of the output value $y$ can be inferred using the following equations:

$$y = \frac{\sum_{k=1}^{k} W^k \mu^k}{\sum_{k=1}^{k} W^k} \tag{6.5}$$

where

$$W^k = \prod_j A_j^k (X_j) \tag{6.6}$$

The symbol $A_j^k (x_j)$ means the membership function of fuzzy variables, such as $A_{pir}^k (x_{pir})$, $A_{chair}^k (x_{chair})$ ,$A_{pc}^k (x_{pc})$ , $A_{working-time}^k (x_{working-time})$, $A_{time-lapse}^k (x_{time-lapse})$ and $A_{start-time}^k (x_{start-time})$. The symbol $k$ is the number of fuzzy rule, $W^k$ is the

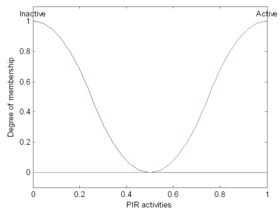
Table 6.3: Fuzzy symbols in consequent part and output value $\mu^k$

| Consequent Part | Meaning of User Activity | Value of $\mu^k$ |
|---|---|---|
| H | At Home or going home | 0 |
| OD | Out of office duties | 1 |
| L | Lunch | 2 |
| S | Short break | 3 |
| NC | Not use computer | 4 |
| C | Using computer | 5 |

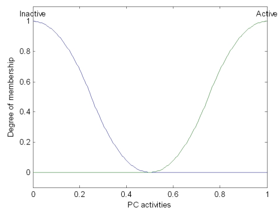
degree of confidence of rule $R_k$, $\mu^k$ is the output value from the rule $R_k$ and $y$ is given by the weighted average of all $R_k$ output. Table 6.3 shows the output value $\mu^k$ from the rule $R_k$ for consequence part with inference output value $y\,(0 \leq y \leq 5)$.

### 6.5.1 Fuzzy Activity Recognition

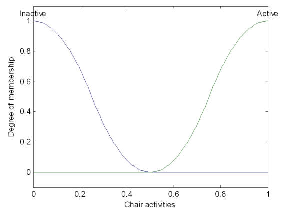
In order to deal with the temporal data of a worker's activity, a fuzzy activity recognition model is constructed. In the first step of to create the FIS, each input variable is fuzzified using linguistic values. The next step is to calculate the fuzzy output and defuzzify it by the calculating the centroid. The output is determined by the range from output Membership Function ($\mu$F) corresponding to each of the six possible events of the worker's activity.




(a) $\mu$F of PIR




(b) $\mu$F of PC

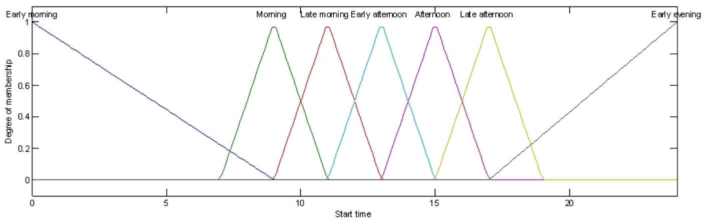



(c) $\mu$F of any binary sensor

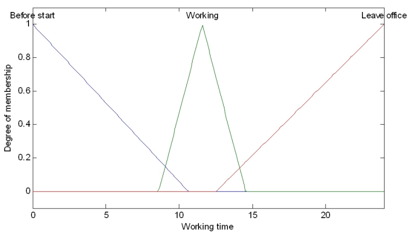
Figure 6.8: The membership functions for active and inactive normalised input values for sensory variables a) PIR, b) PC, and c) any binary sensor.

The $\mu$F of input variables for sensors is defined as shown in Figure 6.8. In fact, the chair's sensor data is used to backup the PIR sensor to identify either a worker in office or out of office.

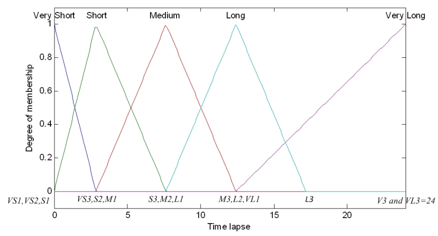
A worker's activity is dynamic and always changing. Therefore, it is important to identify the user's activity based on time. The seven fuzzy partitions for start



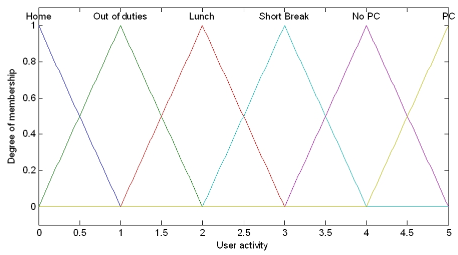

(a) Start time




(b) Working time




(c) Time lapse




(d) User activity

Figure 6.9: The membership functions for a) Start time, b) Working time, c) Time lapse, and d) User activity.

time are shown in Figure 6.9-a. Based on office occupancy data, the working time (start and finish work a day) of a user are determined using a statistical calculation. Figure 6.9-b shows $\mu$Fs of working time for a worker. Time lapse has five $\mu$Fs , and the coordinates of each $\mu$F for five normalized input values of time lapse are shown in Figure 6.9-c. Meanwhile, Figure 6.9-d shows the membership functions of six events of a worker's activity.

The final FIS is created using Fuzzy Toolbox of MATLAB software package. Figure 6.10) show a schematic diagram of the the inference system for an office

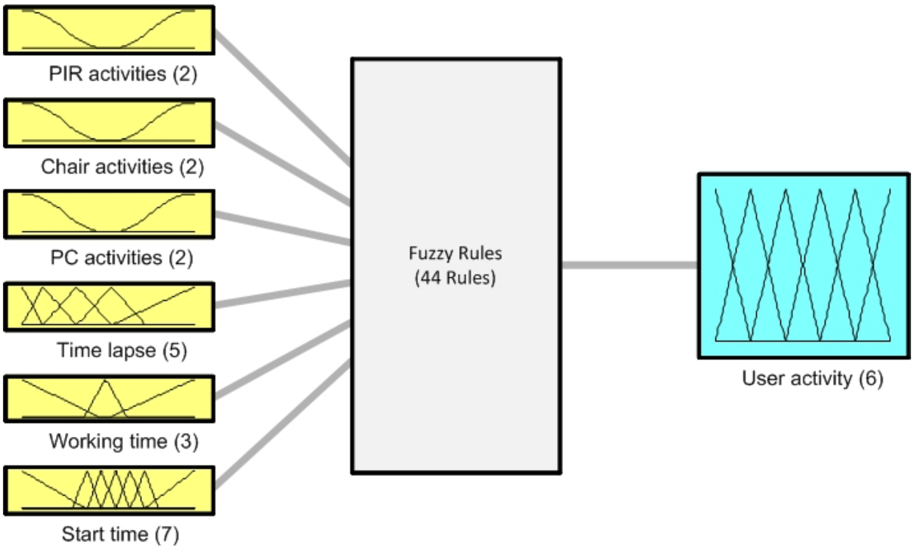


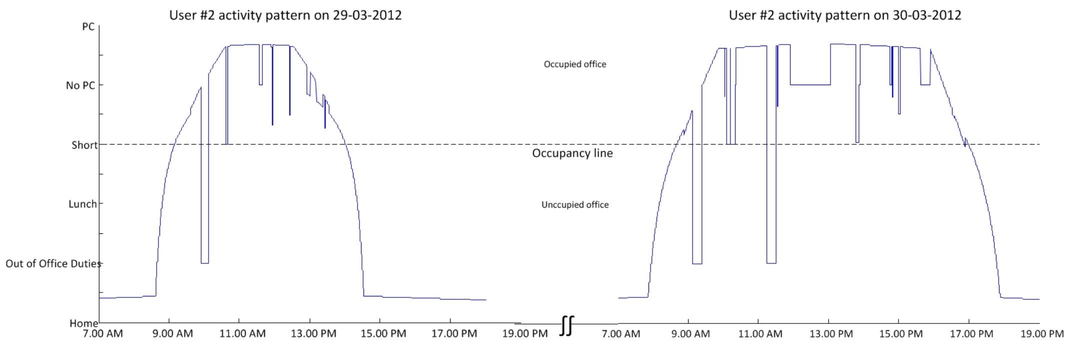

Figure 6.10: Block diagram of a fuzzy activity recognition system for a single office worker.

worker in intelligent office environments. This model consists of six fuzzy input variables, one fuzzy output variable and forty-four fuzzy rules.

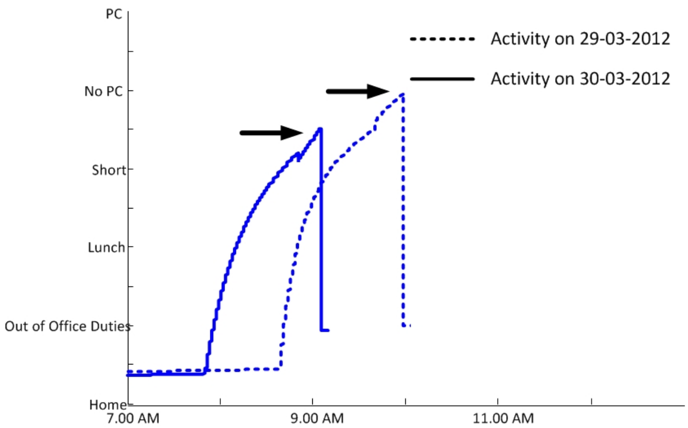
### 6.5.2 Implementation and Results

The Activity of Daily Working (ADW) from data set D5 of User #2 for two different days are used. The results obtained from fuzzy activity recognition are shown in Figure 6.11-a. The graphs show that the patterns of fuzzy user activity are different in shape, compared to the output pattern of an event-driven model. The shape is different because the fuzzy sets in the defuzzification process converted a fuzzy quantity into crisp output values. In the defuzzification process, crisp output values are computed for the fuzzy output variable by aggregating consequence of all activated rules.

Figure 6.11-b compares in detail two curves belonging to the two different days of a user's activities. The fuzzy shapes represent performance of an activity




(a) A fuzzy user activity patterns




(b) The fuzzy curves of user activities

Figure 6.11: Fuzzy activity patterns a) comparing activities for two different days, b) Similar activities in slightly different time.

based on time. As indicated in Figure 6.11-b, arrows show the fuzzy levels of a user's activities are different, resulting from a start time of arrival at the office of a user on day B that is executed earlier (i.e., the user's arrival at the office) than the start working time on day A.

## 6.6 Discussion

This chapter has given an account of enhanced user profiling using fuzzy characteristic matrix. A fuzzy characteristic matrix is to summarise the activities of an office worker based on data collection from sensor network. This technique is undertaken to design fuzzy ambient intelligence for intelligent office environments. A user activity is clustered into fuzzy partitions based on start time and duration. The experimental results in Figure 6.4 shows that activities of a user can be presented in matrix format. The size of circle presented the likelihood of the activity occurrence for specific time, duration and day. From these results, it can be concluded that a fuzzy characteristic matrix enhances our understanding of activities and provides additional information of individual user profiling.

The computational intelligence and pervasive sensing technology offered opportunities to design ambient intelligence for office environment. To optimise the energy consumption of computer and lighting in the office, ADW of an academic office worker must be identified from start to finish. The activity recognition using an event-driven model and fuzzy inference system has the capability to identify the activity of an office worker whether this activity is in the office or out of the office during a typical working day. With this capability, a user activity can be tracked and classified into six categories such as use PC/lighting, not use PC/lighting, short break, lunch break, out of office duties and home. The experiments shows that the proposed activity recognition models are able to identify six categories of user's activity with accuracy of more than 90%. The results obtained from the output of activity recognition are shown in Figure 6.7, it is apparent from the graphs that, unnecessary energy usage of computer and lighting are detected when a user leaves the office. The experimental results shown in Table 6.2 indicated that power consumption by a user for a day that is identified as unnecessary reached 0.099 KWh of computer usage and 0.087 KWh of lighting.

In order to deal with the flexibility and tolerance of activity recognition with user behaviour, a fuzzy inference system is developed which can identify user activities. The findings in this chapter have important implications for developing an office environmental control to archive energy efficiency that responds to a worker behaviour.

# Chapter 7

# Conclusions and Future Works

## 7.1 Summary

This research presents a novel contribution to the development of intelligent office environments. It was undertaken to investigate the ambient and computational intelligence in a smart office environment which can be applied to discover the Activities of Daily Working (ADW) for an office worker, measure the ambient conditions and observe office conditions. The results of this research indicate the pattern of workers' ADW, and the ambience and energy of the office. These variables are the most important roles for the corresponding energy usage behaviour and for constructing the individual user profile. Based on these results, the user profile can enhance our understanding of improved environmental controls in the intelligent office. In addition to the construction of a user profile, the information that is collected based on the ADW of the worker, ambience and energy of the office become important practical variables in data analysis and mining to identify different user profiles. The proposed individual user profile of an office worker comprises routine activities, consistency of office usage and the user's ambient condition preferences. The individual user profiling can be used to automatically adjust the office environment in line with the individual users profile, thus optimising the energy used and improving the office worker's comfort and productivity.

The aim of this research is to profile user activities in office environments in order to improve energy usage and user satisfaction with their working conditions. The real user's activities and ambient conditions in an office environment are monitored using non-intrusive sensors (magnetic door/window sensors, light dependent resistors, Passive Infra-red Motion Sensors, chair pressure pads, temperature and humidity) and a pc monitoring application over a period of time. Therefore, the research was conducted to develop a data collection system, apply the appropriate data mining techniques to construct the individual profile signature representing the individual ADW, and improve to provide a better summary and identification of the worker's behaviour.

Initially, the research was directed to investigate the requirements for the development of an appropriate monitoring system and the suitability of a Wireless Sensor Network (WSN) to measure all parameters with minimum interference with the worker's activities. The data was collected from different office workers using WSN in the real environments of single offices representing the working activities of academic staff. The collected data was gathered into a central database and converted into another format to make it more understandable with regard to the worker's ADW and office environment conditions. The collected data on activities from different single office environments was presented in a binary format, and the ambient data was presented in an analog format. The different sensor measurements were combined and presented challenges for data analysis. Therefore, the challenges of integrating different types of sensors and data representation have been widely investigated. Based on the final data format, the data was presented in event data format, start time-duration format and database format which included date, time, sensor ID, and number of offices/users. The empirical results from these formats provided the capability to reduce the capacity of the database to record data for longer periods of monitoring, and support in applying analysis techniques for further research purposes.

Secondly, using the low level information gathered from the office environment, the data based on ADW and ambient conditions were analysed using statistical analysis, and mined using distance measures. In an attempt to construct the individual user profiles, different techniques including Approximate Entropy (ApEn), Cronbach's $\alpha$ and different similarity measures were employed to quantify the worker's behaviour. The results demonstrated that the chaos level and consistency of a users activities can be determined over a period of time by ApEn and Cronbach's $\alpha$ analysis. ApEn and Cronbach's $\alpha$ analysis provided additional characteristics of workers for the identification of different user profiles. Also, the routine arrival and departure times from the office and the user's thermal comfort level were investigated as a worker's preferences. In addition to pattern extraction, other statistical methods such as moving average, duration average and total number of changes were used to compare different worker's activities over a day/week/month to accurately understand the behavioural pattern of a worker. Different similarity measures were employed to investigate a similarity mining pattern in office worker behaviour. Both linear and non-linear similarity measures, including Dynamic Time Warping (DTW), were investigated to generate the classification of user patterns and subsequently to categorise activities for subsequent energy and comfort optimisation. The integration between DTW and an algorithm of behavioural changes identification was proposed to identify the behavioural changes of a worker over a period of time. Based on the similarity pattern results, the ADW of an office worker can be discovered to distinguish the regular behaviour that naturally occurs in a worker's work routine.

Finally, in order to enhance user profiling, the challenges of employing soft computing techniques to create a more specific user characteristic over time were widely investigated. A new characteristic presentation of a worker's activities using a fuzzy characteristic matrix was proposed. The fuzzy characteristic matrix was used to present an activity of office workers for a specific start time and the duration for a day of the week. In the university's office environment, office users in this study could be seen in instances of lecturing and management schedules that had them displaying intermittent office use. Therefore, different soft computing techniques, including an event-driven and a Fuzzy Inference System (FIS), were investigated in order to recognise a worker's activities during times when the office is occupied and unoccupied for a workday. The experimental results demonstrated that the accuracy of the activity recognition models was more than 90 %, and the models were capable of recognising the ADW of a worker and were able to classify them into six categories (home, lunch, short break, out of office duties, not use computer/lighting and use computer/lighting). The significant results of these models for activity recognition were highlighted by their capability to apply sensory signals for the identification of patterns for office workers' behaviour and office power usage.

## 7.2 Concluding Remarks

This thesis attempts to generate a classification of user patterns and subsequently to categorise activities for subsequent energy and comfort optimisation. Conclusions for the proposed solutions of the project are explained below:

### 7.2.1 Data Acquisition and Collection in Intelligent Office Environments

The focus of the research was the office environment. The main part of the investigation was to understand an office workers' behaviour for improved environmental control in the intelligent office environment. The office workers' comfort, office energy efficiency and worker productivity in an office environment can depend on the temperature, ventilation and lighting. Therefore, the intelligent office environment was proposed as shown in Figure 3.1 where the three important aspects such as the ADW of a worker, the ambient conditions and energy usage of the office were monitored and investigated. The acquisition and collection of data were carried out by WSN and consisted of sensor-based activity recognition; sensor based environmental monitoring and a central database. The experimental setup was conducted at the Computing and Informatics Building, Clifton Campus, Nottingham Trent University, Nottingham, UK. The results of the collected data provided the information for analysing and recognising the patterns of office workers' behaviour, ambient conditions and energy usage.

### 7.2.2 Data Representation and Visualization

The experimental office was set up as discussed in Chapter 3. It collected longer logging sequences from different offices with the academic staff volunteers. Different data representation techniques such as binary time series, analog time series and start-time and duration techniques were applied to represent the ADW of office workers, ambient conditions and power usage. The results of this data representation were successful in converting the large binary/analog sensors data sets to a simple and more meaningful format. Therefore, the formats produced by these data representation techniques can assist to analyse and identify patterns of ADW, ambient conditions and power usage. Also, the data collected from the real environment was used to guide the development of an office simulator. The simulator was created to generate large sensory signals which represent different activities and occupancy of the office environment. It was shown that the simulator signals provided a test bed for the various algorithms which were needed to assess and compare, and helped create an increasingly more sophisticated environment.

### 7.2.3 Individual User Profiling Extraction

The research was comprehensive in discovering issues related to the user profiling extraction and construction for office workers in an intelligent office environment. The ADW of a worker, the ambient temperature and patterns of office use and energy usage, the related aspects of office energy and comfort optimisation were addressed properly in this research. The individual user profile was defined as being associated with the user's work habits and his/her preferences such as user routine activities, consistency of each user's activity and user thermal comfort. Therefore, this research particularly investigated the effectiveness of an ApEn algorithm, a coefficient of internal consistency and both linear and non-linear similarity measures to identify different individual user profiles in an intelligent office environment.

The chaos levels of the workers' activities over time (day, week or month) were computed using an ApEn algorithm. Based on the results of the chaotic worker activity obtained from the ApEn, it can be concluded that each office worker has different routines of ADW. It can also be used to identify different chaos levels and determine the predictable levels of the worker's behaviour. The duration movement pattern of the ADW was used as a parameter to measure the internal consistency, while the Cronbach's $\alpha$ was used to determine the level of consistency for the worker's activities based on a range of 0 to 1, where, the value of $\alpha$ is considered to be satisfactory consistency at $\alpha > 0.5$, while alpha values above 0.7 indicate extremely high consistency [158]. In addition to identifying the different user profiles, the moving average was used to determine the thermal comfort of each worker. As discussed in Chapter 5, the real data ($D4$, $D5$ and $D6$) was used to generate a classification of user profiles as shown in Table 5.5. The experimental results showed that the extraction techniques (see in Table 5.6) succeeded in producing and identifying different individual user profiles in representing the ADW of workers in the intelligent office environment.

To identify similarities or dissimilarities between different worker's ADW and also to compare the user's behaviour across different days/weeks it is important to investigate different measures which could accurately represent the relationship between days and users. In this research, some linear similarity measures have been investigated along with dynamic time warping to measure the similarity of activities in office environments. The results of this investigation show that the degree of normal and abnormal activities of the user between different days and different users can be measured using the Jaccard similarity measure, Gower and Legendre, Hamming Distance and dynamic time warping. In addition, the similarities or dissimilarities of the workers' activities are related to the energy use behaviour of the worker while working in an office environment. The Pearson correlation coefficient ($r$) was used to measure the degree of the relationship between different similarities of the worker's activities. Based on the results, the empirical findings provided the relationship model for obtaining estimates of the user's behaviour based on energy usage. It was shown that the relationship models (see in Figure 5.12) based on daily/weekly similarities observations can assist the environmental control to learn the energy use behaviour of a worker while working in an office environment.

The research was undertaken to generate a classification of user patterns, categorise activities for subsequent energy optimisation and identify a worker's behavioural change over time. Different techniques, including hierarchical clustering, Principal Component Analysis and development algorithm of behaviour changes identification were widely investigated. As discussed in Chapter 5, the statistical experimental results based on the real data had clustered similarities between different worker's activities (see Figure 5.15) and detected the workers behaviour based on time (see Figure 5.18-c). These techniques can be used to generate a classification of an office worker pattern and subsequently categorise activities for subsequent energy and comfort optimisation.

### 7.2.4 Enhance Extraction and Identification of ADW

In this study, methods of describing the behaviour of an office user have been explored using fuzzy values. It appears to effectively summarise the ADW of an office worker more specifically from the start time and for the duration of that day of the week. The profile matrix is presented as fuzzy values which indicate the likelihood of a sequence of activities. The findings from this technique will serve as a basis for future studies for the control of office environments based on a simple profiling from fuzzy user characteristics.

The idea of ADW identification in single office environments was extended by activity recognition. The identification of behaviour over time would also be beneficial for identifying significant changes in office power usage, for example, leaving the office to go to the toilet, lecturing, annual leave, etc. The activity recognition models, which are event-driven and use a FIS, are proposed for the recognition and categorisation of the ADW of a worker. The proposed models are capable of identifying the worker's activities during times when the office is occupied and unoccupied for a workday. The activity recognition models were used to uncover issues related to identifying the activities of a worker for subsequent energy and comfort optimisation. Based on the experimental results, it was proven that an office worker's activities in a workday can be classified into six categories (home, lunch, short break, out of office duties, not use computer/lighting and use computer/lighting) with an accuracy of more than 90 %.

## 7.3 Directions for Future Work

Further investigations could be carried out for future work as listed below:

- There are a number of important parameters which need to be expanded to generate an individual user profile for comfort optimisation rather than only monitor thermal comfort based on office temperature and humidity. It would be interesting to investigate the sensing of human effects (e.g. anxiety, stress, blood pressure and breath rate) on office users in an office environment. Physiological sensors [171] such as the Blood Volume Pulse (BVP) to detect the blood pressure, the Galvanic Skin Response (GSR) to detect anxiety due to increased activity in the sweat glands, the Electromyography (EMG), and the respiration sensor for diaphragmatic breathing monitoring offer unprecedented opportunities to provide data continuously without interfering with the user's daily activities. More information on the physical state or behaviour of an office user would help to extract a user profile which can be used to assist the environmental control to automatically adjust the office environment in line with the individual user profile, thus optimising energy use and improving the office workers comfort and productivity.

- This study will enhance efforts for future studies in collecting the ADW data from multiple office workers using a proposed data acquisition mechanism and representing the multiple users profile in office building environments. Also, further research into the prediction of real office users' activities, ambient conditions and office power usage would be of great help in improving the intelligent office system. With this capability, an environmental control system can be applied to auto-adjust the energy and thermal comfort levels based on the users' preferences.

- Considerably more work will be required in order to provide efficient monitoring of office energy consumption and control of the office environment. The current study has covered the development of data collection to monitor the user's ADW, user activity recognition, and the generation of user profiles which are related to office energy and comfort optimisation. Future studies will have to be directed toward the effective integration of the data collection system, activity recognition system and individual user profiling to the office environment control agents. This new concept may be introduced into smart offices and can improve energy efficiency and comfort in line with individual user profiles.

- It will be interesting to extend the work to develop Data Visualisation Systems to simulate and visualise the movement patterns of different office users in different office environments. The model design will include the movement patterns of the user, the office layout, the duration of time spent in the office and the use of the computer/light/heater, and allow comparisons with the previous user's behaviour pattern or with other users behaviour patterns in an office environment.

- The activity recognition models developed in this work could be extended to detect human behaviour based on real time. For future planning, some predictive algorithms could be investigated to recognize the ADW and to allow a combination with an activity recognition system to develop a personal assistance system where the system works in the same way as a personal secretary to assist office users to manage their work activities or to provide information to others, such as the current or future user's work activity in an Enquiry System.

# Appendix A - Current Products for Office Building Monitoring

Based on our survey, list of companies which provide necessary hardware and software for office building monitoring are presented in this appendix. Table 1 lists hardware products for office building monitoring applications and Table 2 lists software products for office building monitoring applications.

Table 1: Hardware products for office building monitoring applications.

| Company | Technology | Application |
|---|---|---|
| Human Recognition System www.hrsid.com | HRS which is integrating biometric and behavioural analytic technology | The system that can verify, monitor, report and alert on people movement to increase security and improve operational effort. |
| Trend Control System www.trendcontrols.com | IQ ASSURED is structured program for Building Energy Management System (BEMS). | To review, monitor, optimise and demonstrate of building energy consumption. |
| Resource Data Management Ltd. www.resourcedm.com | Building Management System (BMS) | To reduce energy consumption and cost. |
| SKC Inc. www.skcinc.com | QUESTempo 36 is Datalogging Monitor includes sensors and able to link with Detection Management Software (DMS). | To retrieve, report, share, store data collected for indoor thermal building comfort monitoring. |
| LumaSense technologies Inc. www.lumasenseinc.com | Thermal Comfort Datalogger-INNOVA is connected to a PC and work with the application software INNOVA 7701. | Measuring of all physical parameters necessary to evaluate heat stress and thermal comfort. |
| OKI www.oki.com | Ultra-sensitive Human-detecting Sensor Technology | Currently seeking to apply this technology to areas ranging from security to the monitoring of elderly or people requiring long-term care. |
| OWL www.theowl.com | The OWL Z-Smart In Home Display uses the Zigbee Smart Energy standard. | To control and monitor electricity usage. |
| SourceSecurity.com www.sourcesecurity.com | Sight-sensor thermal cameras detect intrusions and auto-steer Pantiltzoom (PTZ) using Sight-tracker | The solution is a hands-free, auto tracking PTZ controller that automatically zooms a PTZ to follow an intruder in real time. |
| Schneider Electric www.powerlogic.com | Using PowerLogic ION EEM | To extends the benefit of existing energy-related data resources. |

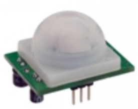
Table 2: Software products for office building monitoring applications.

| Company | Software Technology | Application |
|---|---|---|
| BizEE Software Ltd. www.energylens.com | Energy Lens is Energy Management Software (EMS) | Tool for charting and analysing energy consumption. |
| eSight Energy Group Ltd. www.esightenergy.com | Ltd.www.esightenergy.com eSight Energys eSight M&T or EMS (Monitoring and Targeting or Energy Management System) | To collate, manage and monitor all aspects of energy and building-related data in a single system |
| Parc Tecnologic Barcelona www.dexmatech.com | DEXCell Energy Manager is an online software | To monitoring, analysis, alarming, reporting and awareness, compatible with all types of meters (electricity, heat,temperature, humidity, gas and water). |
| Hunter Engineering Services Ltd.:www.meterology.co.uk | METERology is an engineered solution for monitoring all energy meters through the web application | To profile, analysis, summarise and compare the building's energy. |
| Aussie Home Energy aussiehomeenergy.com.au | Web Portals For Home Electricity Monitors. | To monitor solar PV and power usage. |
| Open source openenergymonitor.org | OpenEnergyMonitor is open-source energy monitoring tools | To visualise, log and process energy data with a powerful open source web. |
| EG Energy Controls www.egenergy.com | The Energy Surveillance System (ESS) is a energy monitoring system that is web based. | It is allows you to compare energy usage by equipment type, store location, fuel type or time of day. |
| Carbonetix www.carbonrealtime.com.au | Monitor Energy Use With Carbon Real-Time Energy Monitoring Software | To monitor energy use of real-time for electric usage and ambient temperature. |
| Energy Auditing Agency www.teamenergy.com | The TEAM Sigma software is energy monitoring and targeting | Database management, sharing information, analysis and etc.. |

# Appendix B - Sensor Types

Description of the sensory devices employed in monitoring the experimental environment are illustrated in this appendix.

1. Sensor-based activity recognition

| Sensor | Image | Type | Data type | Application |
|---|---|---|---|---|
| Passive Infrared |  | Motion detector infrared | Digital | Motion detection. The PIRs is active when detecting people movement. |
| Door / Window Contact |  | Magnetic Switch | Analog | To detect the open and closed states of door and window. |
| Pressure Pad |  | Pressure | Analog | To identify the occupied and unoccupied states of chair occupancy. |
| Light Dependent Resistor |  | Photo-resistor | Analog | To detect the ON and OFF states of office's lights. |
| Inductive Current |  | Current Transformer | Analog | To detect the current level of power cable for computer. |

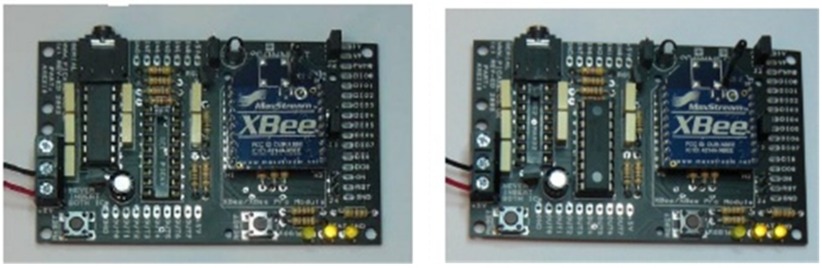
2. Sensor-based environmental monitoring

| Sensor | Image | Type | Data type | Application |
|---|---|---|---|---|
| DS18B20 digital thermometer |  | Temperature | Digital | To measure the environment temperature. |
| HIH 4000 Humidity |  | Humidity | Analog | To measure the environment relative humidity. |
| Light Dependent Resistor |  | Photo-resistor | Analog | To determine the intensity of ambient light level. |

# Appendix C - Sensor Node Devices

In order to integrate the sensor systems and the wireless communication devices, PICAXE Connect AXE210 development board is used. Figure 1-a shows an assembled board of AXE210. It can transmit and retrieve data through a XBee communication modules. Figure 1-b shows that sensor node with XBee wireless as PC interface is use transceiver SP3232E IC and a cable be connected between a PC and jack connector on the board. The function of SP3232E IC is converts signals from a serial port to signals suitable for use in TTL compatible digital logic circuits.

Each board is controlled by a PICAXE micro-controller unit. Pinout configuration of PICAXE micro-controller is shown in Figure 2-a. The main advantage is that PICAXE-18X can be program easily using free software BASIC programming editor (`http://www.picaxe.com/Software`) where it can use data out and data input to respond to the reception or sending of data, such as wake up XBee from sleep mode to transmit an information if received different value of data compared previous data. PICAXE Programming Editor is BASIC programming which is then downloaded into a PICAXE IC. The editor allows for creation of



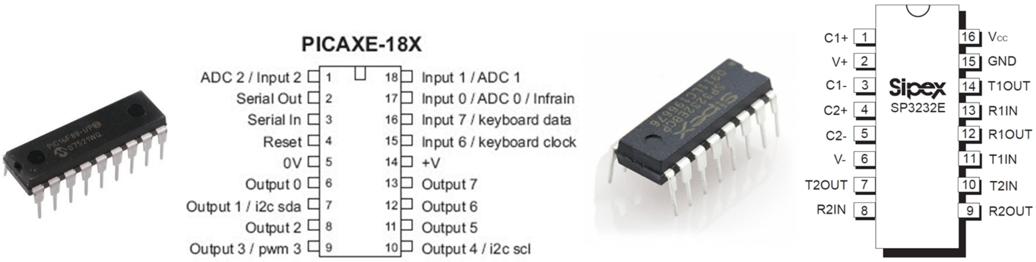
PC interface XBee wireless network (IC MAX3232 is used)

(a)

Picaxe interface XBee wireless network (IC PICAXE-18X is used)

(b)

Figure 1: PICAXE Connect AXE210 a) Sensor board with XBee wireless for PC interface, b) Sensor board with XBee wireless for PICAXE interface




(a) PICAXE-18X

(b) SP3232E

Figure 2: Pictures and pinout configurations for a) PICAXE-18X and b) SP3232E

programs in a variety of ways inclusive of both way such flowchart and BASIC programs.

Pinout configuration of SP3232E is shown in Figure 2-b as a dual driver or receiver and conventionally to convert the RX, TX, CTS and RTS signals. Therefore, the Xbee module is connected directly to serial port of the computer via the SP3232E IC. While sensor node with XBee wireless as PICAXE interface is use PICAXE IC and can be used for all data acquisition.

The PICAXE micro-controller contains 2048 Bytes of programmable flash memory (up to 1800 lines of program), 512 Byte of static Random Access Memory (RAM), and new lower down to 1.8 V operation makes ideal for use with 3

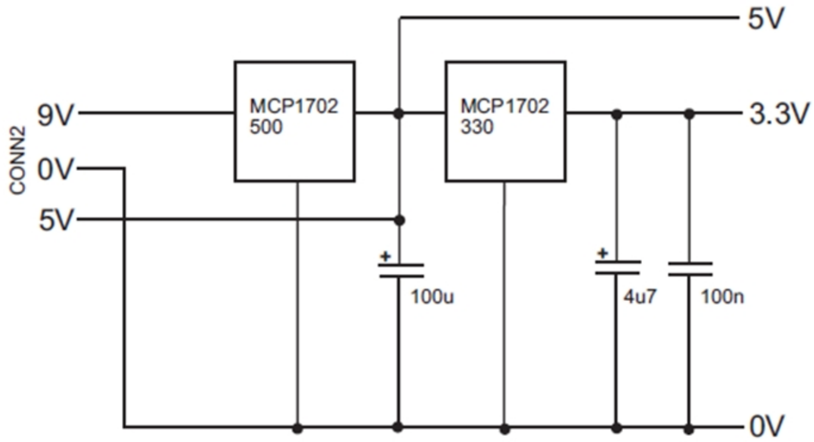



Figure 3: Circuit diagram of AXE210 power circuit

V battery i.e. save the cost (http://www.picaxe.com/docs/picaxem2.pdf). Actually, the PICAXE micro-controller ICs are currently being designed in six different sizes, number of pins and come with two series (M2 and X2 series). Table 1 shows the compared of the PICAXE's micro-controller ICs.

The PICAXE AXE210 is designed to operate with a voltage of +5 V and +9 V regulated. Two voltage regulators on the board such as a MCP1702-500 regulator provides the 5 V from power supply to ICs on 1702-330 provide 3.3 V to XBee module. The three batteries AA with 4.5 V enough to be used to

Table 1: Summary of the various PICAXE ICs.

| PICAXE Type | IC Size | Memory (Lines) | I/O Pins | Outputs | Inputs |
|---|---|---|---|---|---|
| PICAXE-08 | 8 | 40-110 | 5 | 1-4 | 1-4 |
| PICAXE-08M | 8 | 80-220 | 5 | 1-4 | 1-4 |
| PICAXE-14M | 14 | 80-220 | 13 | 5 | 6 |
| PICAXE-18M | 18 | 80-220 | 13 | 8 | 5 |
| PICAXE-18X | 18 | 600-1800 | 14 | 9 | 5 |
| PICAXE-20M | 20 | 80-220 | 18 | 8 | 8 |
| PICAXE-20X2 | 20 | 1000-3200 | 18 | 1-17 | 1-17 |

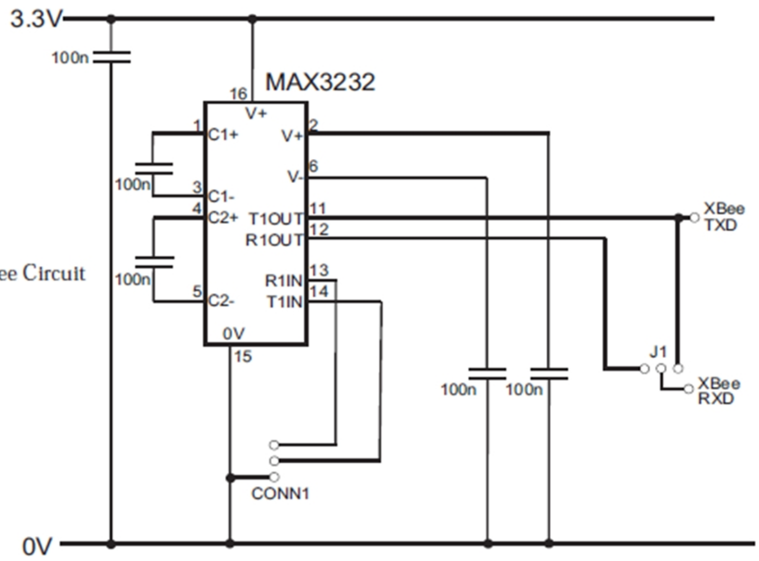



Figure 4: Circuit diagram of SP3232E IC to XBee module connection for PC interface

supply power to PICAXE IC, SP3232E IC and XBee. The schematic diagram of AXE210 board power circuit is shown in the Figure 3.

The PICAXE AXE210 connect board was designed to operate in two different modes and the schematic for the respective mode are explained below:

1. PC interface (SP3232E IC inserted): The second mode of operation is where an interface between the PICAXE AXE210 board and a computer is employed. Here, a SP3232E IC is put in place of the PICAXE IC and this allows for transmitting and receiving from the computer and also allows for configuration of the XBee module via a direct connection of the computer's serial port. The schematic diagram of SP3232E IC to XBee module connection is shown in Figure 4.

2. PICAXE interface (PICAXE-18X IC inserted): The first mode is where it creates an interface between the PICAXE micro-controller and the Xbee

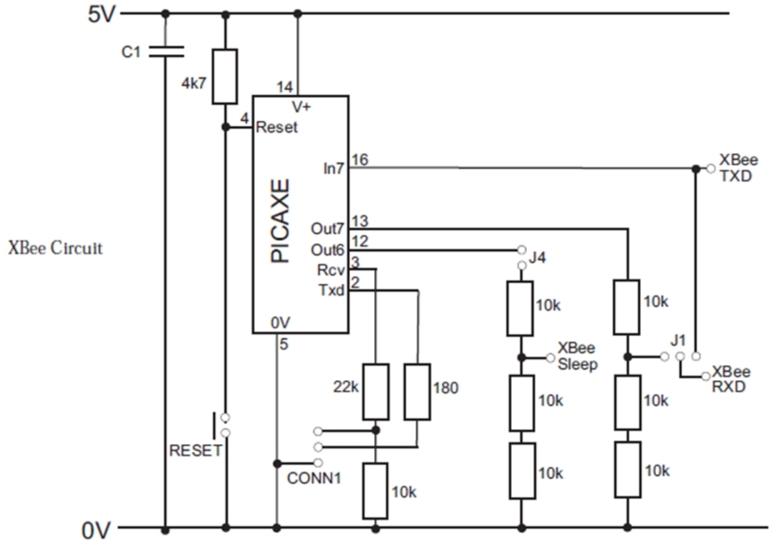
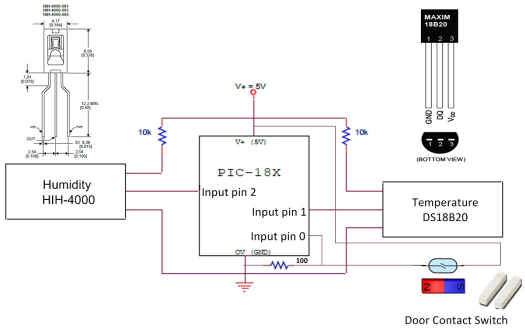


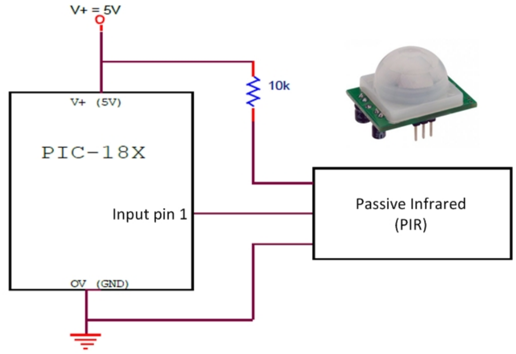

Figure 5: Circuit diagram of PICAXE to XBee module connection

module. This design of setup is employed at nodes where sensors have to be connected to. Figure 5 shows the schematic diagram for connection of PICAXCE IC to XBee module. It can be seen from the schematic that a standard set down format is employed during setting up these boards. From the PICAXE IC, it can be seen that output pin 7 is linked connected directly to the XBee transmit pin through jumper J1. Output pin 6 is connected to the XBee sleep pin via jumper 4 and Input pin 7 is connected to the Xbee receive pin. Utilizing the laid down connection standard, it can be seen that data can be transmitted from the Xbee module via output pin 7 and data can be receive by the XBee module via input pin 7.

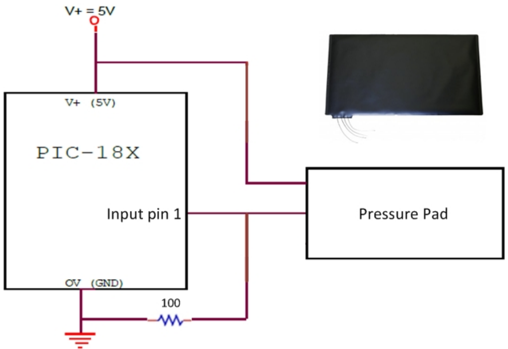
Since the PICAXE IC operates at 5 V, which is higher than the 3.3 V that the XBee module operates in. It can be seen that a simple potential divider arrangement is put in place to interface between the PICAXE IC and XBee module.



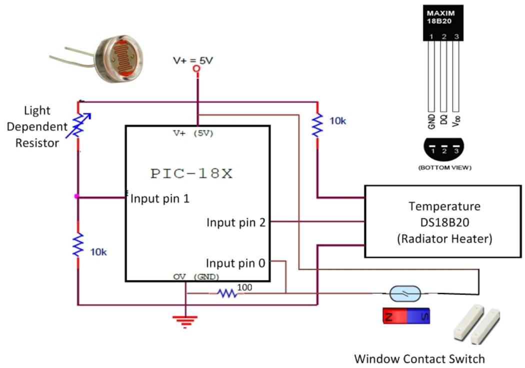

(a) Sensor node #1



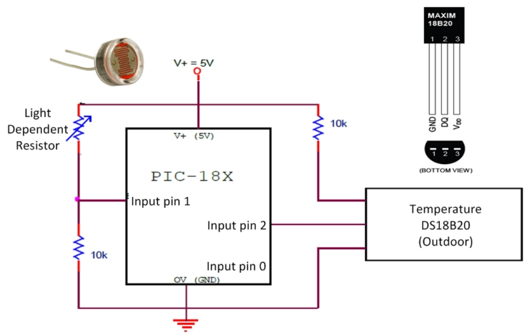

(b) Sensor node #2




(c) Sensor node #3




(d) Sensor node #4




(e) Sensor node #5

Figure 6: The Circuit diagrams of sensors to PICAXE IC connection for a) Sensor node #1, b) Sensor node #2, c)Sensor node #3, d)Sensor node #4 and e)Sensor node #5

The mode PC interface is used for all data acquisition, while the mode PICAXE interface is used as sensor node to collect data from sensor and then send it to coordinator. Although all sensory device do not get directly connected to the PICAXE AXE210 board, a resistor is integrated into the connection between the board and the sensor, it is to act as a form of protection to the sensor. For example, Figure 6 shows that how to connect sensors such as DS18B20 temperature, LDR, HIH 4000 humidity, pressure pad, magnetic switch and PIR to PICAXE IC on an AXE210 board.

# Appendix D - Statistical Analysis of Users' Behaviour

Detailed statistical analysis of users' behaviour are presented in this appendix.

Table 1: Summary of statistical analysis results of data set $D4$. They are recorded from 09 Jan 2012 to 01 April 2012.

| Time in hours | | Weeks | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Total /week | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| **Door** | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| | **Duration** | 7.552 | 7.574 | 5.850 | 4.135 | 3.381 | 1.771 | 4.117 | 6.055 | 2.700 | 2.845 | 5.266 | 3.958 |
| | Max | 4.587 | 2.576 | 2.011 | 1.602 | 2.281 | 1.214 | 1.787 | 2.530 | 1.257 | 1.323 | 2.570 | 1.865 |
| | Min | 0.220 | 0.558 | 0.469 | 0.206 | 0.193 | 0.041 | 0.034 | 0.092 | 0.176 | 0.008 | 0.250 | 0.005 |
| | **Changes** | 139 | 156 | 172 | 194 | 233 | 174 | 139 | 189 | 178 | 100 | 110 | 55 |
| | Max | 47 | 48 | 46 | 48 | 70 | 52 | 48 | 47 | 40 | 35 | 29 | 19 |
| | Min | 5 | 10 | 22 | 22 | 28 | 12 | 12 | 32 | 11 | 2 | 24 | 2 |
| | **Mean** | 0.066 | 0.037 | 0.025 | 0.019 | 0.014 | 0.006 | 0.014 | 0.023 | 0.027 | 0.017 | 0.027 | 0.057 |
| | Max | 0.304 | 0.083 | 0.047 | 0.073 | 0.074 | 0.024 | 0.037 | 0.070 | 0.113 | 0.060 | 0.092 | 0.266 |
| | Min | 0.012 | 0.019 | 0.013 | 0.005 | 0.004 | 0.003 | 0.003 | 0.003 | 0.006 | 0.004 | 0.010 | 0.003 |
| **Occupancy** | **Duration** | 14.26 | 32.41 | 8.67 | 27.73 | 11.57 | 11.89 | 3.87 | 22.30 | 23.28 | 31.78 | 20.84 | 7.73 |
| | Max | 7.32 | 8.23 | 8.58 | 8.22 | 3.62 | 5.07 | 2.27 | 7.52 | 5.14 | 13.39 | 10.85 | 3.61 |
| | Min | 0.00 | 2.90 | 0.01 | 1.90 | 0.70 | 0.02 | 0.03 | 1.78 | 4.06 | 0.02 | 0.28 | 0.03 |
| | **Changes** | 6 | 14 | 1 | 29 | 66 | 51 | 24 | 72 | 105 | 84 | 117 | 62 |
| | Max | 3 | 5 | 1 | 13 | 19 | 21 | 12 | 20 | 39 | 42 | 40 | 22 |
| | Min | 1 | 1 | 1 | 1 | 6 | 5 | 1 | 10 | 2 | 6 | 4 | 10 |
| | **Mean** | 1.375 | 2.215 | 1.226 | 1.905 | 0.109 | 0.115 | 0.052 | 0.248 | 0.193 | 0.379 | 0.096 | 0.052 |
| | Max | 7.324 | 8.076 | 8.584 | 8.217 | 0.278 | 0.398 | 0.175 | 0.745 | 0.517 | 1.322 | 0.404 | 0.164 |
| | Min | 0.054 | 0.573 | 8.584 | 0.127 | 0.077 | 0.077 | 0.024 | 0.103 | 0.018 | 0.113 | 0.019 | 0.055 |
| **Light** | **Duration** | 24.47 | 20.84 | 29.97 | 24.93 | 16.80 | 18.55 | 10.15 | 11.86 | 17.74 | 15.59 | 12.37 | 14.89 |
| | Max | 6.85 | 6.28 | 6.91 | 7.31 | 6.77 | 7.77 | 7.62 | 6.96 | 6.60 | 6.21 | 7.05 | 6.61 |
| | Min | 1.16 | 3.97 | 4.61 | 1.38 | 0.58 | 1.18 | 0.56 | 0.52 | 2.70 | 0.51 | 0.47 | 1.86 |
| | **Changes** | 11 | 11 | 9 | 26 | 6 | 8 | 3 | 6 | 12 | 16 | 19 | 31 |
| | Max | 4 | 4 | 2 | 15 | 2 | 3 | 1 | 3 | 4 | 6 | 8 | 10 |
| | Min | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 |
| | **Mean** | 2.025 | 1.514 | 2.634 | 1.636 | 1.435 | 1.450 | 1.449 | 1.032 | 1.975 | 0.947 | 0.765 | 0.447 |
| | Max | 6.706 | 6.260 | 6.915 | 3.653 | 3.384 | 4.771 | 7.625 | 2.320 | 6.604 | 3.107 | 3.523 | 1.186 |
| | Min | 0.581 | 1.324 | 2.303 | 0.141 | 0.576 | 1.183 | 0.559 | 0.518 | 0.019 | 0.020 | 0.035 | 0.298 |
| **Chair** | **Duration** | 3.96 | 10.01 | 0 | 7.00 | 6.48 | 7.84 | 2.10 | 12.12 | 8.21 | 3.15 | 4.28 | 3.71 |
| | Max | 2.72 | 3.98 | 0 | 2.26 | 1.81 | 3.88 | 1.07 | 4.57 | 2.32 | 1.71 | 1.81 | 1.56 |
| | Min | 0.14 | 0.06 | 0 | 0.34 | 0.28 | 0.34 | 0.02 | 0.70 | 1.37 | 0.06 | 0.32 | 0.34 |
| | **Changes** | 14 | 33 | 0 | 42 | 62 | 43 | 19 | 41 | 62 | 25 | 44 | 30 |
| | Max | 10 | 10 | 0 | 12 | 17 | 15 | 11 | 12 | 24 | 14 | 14 | 11 |
| | Min | 1 | 1 | 0 | 4 | 3 | 4 | 8 | 5 | 6 | 1 | 5 | 3 |
| | **Mean** | 0.200 | 0.197 | 0 | 0.117 | 0.064 | 0.091 | 0.031 | 0.231 | 0.116 | 0.049 | 0.049 | 0.066 |
| | Max | 1.086 | 0.663 | 0 | 0.303 | 0.120 | 0.312 | 0.123 | 0.648 | 0.251 | 0.144 | 0.130 | 0.169 |
| | Min | 0.046 | 0.061 | 0 | 0.063 | 0.061 | 0.083 | 0.092 | 0.097 | 0.059 | 0.033 | 0.033 | 0.036 |
| **Computer** | **Duration** | 10.26 | 14.62 | 8.15 | 10.99 | 10.24 | 11.01 | 4.12 | 34.86 | 25.26 | 30.55 | 23.64 | 8.26 |
| | Max | 5.53 | 9.82 | 8.15 | 4.37 | 3.38 | 3.92 | 2.63 | 11.58 | 7.52 | 14.23 | 9.87 | 4.39 |
| | Min | 4.73 | 0.72 | 8.15 | 0.63 | 0.20 | 1.29 | 1.49 | 1.81 | 3.52 | 4.65 | 1.31 | 0.31 |
| | **Changes** | 11 | 18 | 1 | 16 | 19 | 11 | 6 | 46 | 55 | 16 | 22 | 17 |
| | Max | 9 | 6 | 1 | 5 | 5 | 4 | 4 | 12 | 20 | 6 | 10 | 12 |
| | Min | 2 | 3 | 1 | 1 | 1 | 1 | 2 | 6 | 4 | 2 | 2 | 1 |
| | **Mean** | 0.471 | 0.373 | 1.226 | 0.249 | 0.201 | 0.290 | 0.077 | 0.316 | 0.244 | 1.452 | 0.228 | 0.126 |
| | Max | 2.722 | 1.289 | 8.584 | 0.630 | 0.530 | 0.877 | 0.330 | 0.722 | 0.667 | 6.374 | 1.055 | 0.408 |
| | Min | 0.573 | 0.265 | 8.584 | 0.180 | 0.110 | 0.181 | 0.209 | 0.123 | 0.125 | 0.317 | 0.119 | 0.232 |
| **Window** | **Duration** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Max | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Min | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | **Changes** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Max | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Min | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | **Mean** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Max | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Min | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Table 2: The summary of statistical analysis results of $D5$ recorded based on User #2 activity in an office environment. This data has recorded from 09 Jan 2012 to 01 April 2012.

| | Time in hours | Weeks | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Total /week | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| **Door** | **Duration** | 0.535 | 0.347 | 0.636 | 0.410 | 0.180 | 0.126 | 0.112 | 0.413 | 0.279 | 0.429 | 0.301 | 0.214 |
| | Max | 0.321 | 0.091 | 0.303 | 0.155 | 0.057 | 0.048 | 0.044 | 0.246 | 0.093 | 0.221 | 0.139 | 0.073 |
| | Min | 0.056 | 0.049 | 0.003 | 0.024 | 0.003 | 0.022 | 0.032 | 0.018 | 0.034 | 0.002 | 0.021 | 0.017 |
| | **Changes** | 162 | 197 | 163 | 202 | 119 | 84 | 70 | 118 | 156 | 140 | 146 | 137 |
| | Max | 48 | 43 | 56 | 52 | 39 | 32 | 30 | 37 | 36 | 52 | 43 | 42 |
| | Min | 30 | 32 | 2 | 16 | 2 | 14 | 16 | 14 | 23 | 1 | 16 | 12 |
| | **Mean** | 0.002 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.003 | 0.001 | 0.002 | 0.001 | 0.001 |
| | Max | 0.007 | 0.002 | 0.005 | 0.003 | 0.002 | 0.002 | 0.002 | 0.012 | 0.003 | 0.008 | 0.003 | 0.002 |
| | Min | 0.002 | 0.002 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 |
| **Occupancy** | **Duration** | 18.48 | 36.32 | 21.60 | 26.86 | 23.03 | 16.56 | 13.49 | 47.16 | 63.72 | 49.65 | 39.86 | 21.34 |
| | Max | 6.60 | 9.97 | 6.87 | 9.77 | 7.83 | 6.00 | 6.66 | 12.97 | 17.58 | 14.18 | 12.11 | 6.99 |
| | Min | 0.01 | 4.07 | 0.03 | 0.79 | 0.01 | 0.74 | 1.83 | 6.44 | 8.81 | 0.15 | 3.93 | 1.85 |
| | **Changes** | 71 | 15 | 21 | 150 | 14 | 11 | 14 | 27 | 64 | 34 | 38 | 19 |
| | Max | 33 | 5 | 13 | 58 | 5 | 3 | 6 | 12 | 22 | 10 | 23 | 5 |
| | Min | 1 | 1 | 1 | 1 | 1 | 1 | 3 | 1 | 1 | 1 | 2 | 1 |
| | **Mean** | 0.088 | 0.644 | 0.225 | 0.016 | 0.215 | 0.193 | 0.105 | 0.307 | 0.285 | 0.320 | 0.070 | 0.303 |
| | Max | 0.414 | 1.505 | 1.000 | 0.038 | 0.556 | 0.793 | 0.549 | 0.600 | 1.083 | 1.020 | 0.169 | 0.625 |
| | Min | 0.031 | 0.408 | 0.035 | 0.013 | 0.147 | 0.017 | 0.030 | 0.018 | 0.063 | 0.012 | 0.010 | 0.067 |
| **Light** | **Duration** | 24.42 | 29.90 | 28.85 | 35.49 | 29.78 | 20.91 | 13.83 | 30.43 | 32.54 | 29.83 | 30.04 | 31.99 |
| | Max | 7.86 | 8.33 | 8.15 | 8.39 | 8.51 | 7.11 | 7.70 | 7.69 | 7.85 | 8.19 | 7.47 | 8.24 |
| | Min | 5.08 | 5.84 | 0.44 | 3.33 | 0.29 | 2.16 | 6.13 | 3.55 | 4.75 | 6.74 | 4.32 | 5.31 |
| | **Changes** | 12 | 12 | 13 | 12 | 22 | 11 | 30 | 23 | 33 | 26 | 41 | 37 |
| | Max | 4 | 4 | 6 | 3 | 12 | 4 | 15 | 14 | 13 | 14 | 16 | 14 |
| | Min | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 3 |
| | **Mean** | 1.290 | 1.648 | 2.186 | 3.382 | 2.495 | 1.247 | 1.000 | 2.682 | 1.551 | 2.389 | 3.015 | 0.751 |
| | Max | 3.928 | 4.164 | 8.151 | 7.786 | 8.506 | 2.997 | 3.851 | 7.289 | 3.444 | 8.194 | 7.471 | 1.800 |
| | Min | 1.271 | 1.461 | 0.439 | 1.664 | 0.252 | 0.538 | 0.040 | 0.034 | 0.026 | 0.088 | 0.030 | 0.424 |
| **Chair** | **Duration** | 9.07 | 31.36 | 20.12 | 4.12 | 22.13 | 15.61 | 12.90 | 25.69 | 21.83 | 24.13 | 19.11 | 16.27 |
| | Max | 4.73 | 9.87 | 6.72 | 3.04 | 7.57 | 5.42 | 6.51 | 6.35 | 6.34 | 6.85 | 5.67 | 6.54 |
| | Min | 4.34 | 3.56 | 3.11 | 1.07 | 4.05 | 0.66 | 1.77 | 3.53 | 1.86 | 5.19 | 1.79 | 0.84 |
| | **Changes** | 23 | 47 | 35 | 8 | 35 | 32 | 24 | 50 | 41 | 40 | 37 | 36 |
| | Max | 12 | 15 | 11 | 7 | 12 | 13 | 13 | 17 | 16 | 12 | 11 | 11 |
| | Min | 11 | 6 | 5 | 1 | 5 | 4 | 5 | 6 | 4 | 8 | 5 | 3 |
| | **Mean** | 0.113 | 0.501 | 0.329 | 0.215 | 0.382 | 0.265 | 0.232 | 0.393 | 0.416 | 0.351 | 0.359 | 0.299 |
| | Max | 0.430 | 1.233 | 0.621 | 1.072 | 0.918 | 0.696 | 0.768 | 0.728 | 0.810 | 0.757 | 0.593 | 0.594 |
| | Min | 0.360 | 0.464 | 0.487 | 0.435 | 0.492 | 0.164 | 0.355 | 0.348 | 0.315 | 0.506 | 0.356 | 0.275 |
| **Computer** | **Duration** | 16.45 | 43.05 | 18.72 | 42.05 | 28.09 | 21.18 | 25.46 | 36.86 | 25.21 | 49.55 | 56.06 | 21.92 |
| | Max | 5.93 | 19.05 | 6.62 | 21.44 | 7.81 | 10.12 | 7.69 | 11.34 | 8.25 | 21.12 | 15.51 | 8.89 |
| | Min | 3.08 | 3.94 | 3.10 | 1.94 | 0.95 | 3.08 | 5.48 | 2.19 | 0.99 | 2.98 | 6.74 | 0.34 |
| | **Changes** | 39 | 77 | 50 | 68 | 54 | 41 | 52 | 67 | 51 | 80 | 59 | 36 |
| | Max | 11 | 29 | 24 | 17 | 17 | 18 | 19 | 21 | 17 | 27 | 16 | 12 |
| | Min | 8 | 7 | 5 | 7 | 2 | 4 | 9 | 6 | 2 | 9 | 7 | 5 |
| | **Mean** | 0.201 | 0.347 | 0.222 | 0.393 | 0.346 | 0.310 | 0.252 | 0.345 | 0.318 | 0.397 | 0.660 | 0.298 |
| | Max | 0.476 | 0.608 | 0.559 | 1.597 | 0.661 | 1.149 | 0.517 | 0.845 | 0.599 | 1.307 | 1.532 | 0.648 |
| | Min | 0.232 | 0.375 | 0.226 | 0.193 | 0.317 | 0.233 | 0.280 | 0.301 | 0.300 | 0.225 | 0.382 | 0.343 |
| **Window** | **Duration** | 4.50 | 3.69 | 0.48 | 0.73 | 0.10 | 0.00 | 10.58 | 11.41 | 7.02 | 1.64 | 0.00 | 4.77 |
| | Max | 1.87 | 2.51 | 0.33 | 0.57 | 0.10 | 0.00 | 8.13 | 4.90 | 4.60 | 1.59 | 0.00 | 2.12 |
| | Min | 1.26 | 0.27 | 0.15 | 0.05 | 0.10 | 0.00 | 2.45 | 0.98 | 2.41 | 0.05 | 0.00 | 0.83 |
| | **Changes** | 11 | 11 | 3 | 3 | 1 | 0 | 5 | 22 | 19 | 8 | 0 | 7 |
| | Max | 5 | 7 | 2 | 1 | 1 | 0 | 4 | 15 | 12 | 7 | 0 | 3 |
| | Min | 2 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 7 | 1 | 0 | 2 |
| | **Mean** | 0.192 | 0.177 | 0.058 | 0.105 | 0.014 | 0.000 | 0.640 | 0.766 | 0.104 | 0.040 | 0.000 | 0.297 |
| | Max | 0.626 | 0.836 | 0.331 | 0.571 | 0.097 | 0.000 | 2.448 | 4.049 | 0.383 | 0.227 | 0.000 | 1.060 |
| | Min | 0.342 | 0.130 | 0.077 | 0.045 | 0.097 | 0.000 | 2.032 | 0.244 | 0.345 | 0.052 | 0.000 | 0.413 |

Table 3: The summary of statistical analysis results of *D*6 recorded based on User #4 activity in an office environment. This data has recorded from 09 Jan 2012 to 01 April 2012.

| | Time in hours Total /week | Weeks 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Door** | **Duration** | 0.638 | 0.298 | 0.503 | 0.396 | 0.401 | 0.129 | 0.073 | 0.126 | 0.293 | 0.796 | 0.408 | 0.192 |
| | Max | 0.415 | 0.114 | 0.274 | 0.159 | 0.156 | 0.050 | 0.051 | 0.052 | 0.187 | 0.636 | 0.124 | 0.166 |
| | Min | 0.015 | 0.023 | 0.004 | 0.033 | 0.026 | 0.031 | 0.022 | 0.010 | 0.003 | 0.017 | 0.088 | 0.026 |
| | **Changes** | 163 | 159 | 161 | 181 | 217 | 85 | 28 | 71 | 114 | 132 | 165 | 33 |
| | Max | 50 | 55 | 61 | 56 | 80 | 35 | 14 | 27 | 47 | 32 | 61 | 18 |
| | Min | 10 | 10 | 3 | 20 | 19 | 20 | 14 | 5 | 2 | 10 | 15 | 15 |
| | **Mean** | 0.002 | 0.002 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.004 | 0.002 | 0.002 |
| | Max | 0.010 | 0.004 | 0.004 | 0.003 | 0.003 | 0.002 | 0.004 | 0.002 | 0.004 | 0.020 | 0.006 | 0.011 |
| | Min | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 | 0.002 | 0.002 | 0.001 |
| **Occupancy** | **Duration** | 8.24 | 21.55 | 22.54 | 20.96 | 25.94 | 8.76 | 0.03 | 3.35 | 14.51 | 41.37 | 29.48 | 1.24 |
| | Max | 7.51 | 6.46 | 7.86 | 7.28 | 8.14 | 7.63 | 0.03 | 3.27 | 12.39 | 10.90 | 10.51 | 1.21 |
| | Min | 0.01 | 0.72 | 0.09 | 0.00 | 0.01 | 0.12 | 0.03 | 0.01 | 0.01 | 5.64 | 8.93 | 0.03 |
| | **Changes** | 8 | 22 | 9 | 19 | 22 | 6 | 0 | 22 | 71 | 45 | 30 | 2 |
| | Max | 3 | 8 | 3 | 7 | 9 | 3 | 0 | 19 | 44 | 14 | 14 | 2 |
| | Min | 1 | 2 | 1 | 2 | 2 | 1 | 0 | 1 | 6 | 1 | 4 | 2 |
| | **Mean** | 0.558 | 0.894 | 1.849 | 0.907 | 0.968 | 0.593 | 0 | 0.026 | 0.298 | 1.184 | 0.544 | 0.086 |
| | Max | 3.753 | 3.227 | 7.852 | 3.637 | 4.069 | 3.816 | 0 | 0.165 | 2.059 | 5.362 | 2.223 | 0.605 |
| | Min | 0.013 | 0.354 | 0.010 | 0.285 | 0.690 | 0.013 | 0 | 0.010 | 0.013 | 0.415 | 0.749 | 0.605 |
| **Light** | **Duration** | 19.80 | 17.10 | 15.70 | 15.27 | 18.63 | 9.20 | 3.54 | 5.14 | 8.52 | 15.98 | 14.37 | 1.19 |
| | Max | 7.13 | 5.61 | 5.55 | 5.02 | 5.22 | 4.54 | 2.41 | 2.47 | 4.71 | 4.91 | 5.70 | 1.19 |
| | Min | 0.95 | 1.08 | 1.52 | 1.21 | 0.97 | 2.04 | 1.13 | 0.20 | 1.89 | 1.40 | 0.02 | 1.19 |
| | **Changes** | 23 | 28 | 28 | 31 | 26 | 17 | 6 | 14 | 16 | 29 | 19 | 3 |
| | Max | 8 | 7 | 11 | 8 | 7 | 6 | 4 | 6 | 6 | 10 | 7 | 3 |
| | Min | 2 | 4 | 5 | 3 | 2 | 5 | 2 | 2 | 5 | 3 | 1 | 3 |
| | **Mean** | 0.604 | 0.422 | 0.350 | 0.445 | 0.702 | 0.241 | 0.213 | 0.188 | 0.221 | 0.391 | 0.354 | 0.057 |
| | Max | 1.344 | 0.821 | 0.994 | 1.256 | 2.611 | 0.907 | 1.207 | 0.429 | 0.786 | 0.758 | 1.140 | 0.397 |
| | Min | 0.271 | 0.270 | 0.218 | 0.151 | 0.161 | 0.340 | 0.281 | 0.098 | 0.378 | 0.392 | 0.021 | 0.397 |
| **Chair** | **Duration** | 3.66 | 9.85 | 5.83 | 7.75 | 12.74 | 2.36 | 0 | 1.63 | 0.05 | 11.55 | 10.56 | 0.87 |
| | Max | 3.28 | 3.01 | 2.94 | 2.38 | 4.00 | 1.38 | 0 | 1.63 | 0.04 | 3.93 | 4.37 | 0.87 |
| | Min | 0.38 | 0.71 | 0.78 | 1.42 | 2.34 | 0.98 | 0 | 1.63 | 0.00 | 1.14 | 2.98 | 0.87 |
| | **Changes** | 8 | 29 | 20 | 36 | 42 | 9 | 0 | 10 | 1 | 35 | 28 | 5 |
| | Max | 5 | 10 | 7 | 11 | 15 | 5 | 0 | 10 | 1 | 10 | 12 | 5 |
| | Min | 3 | 2 | 6 | 7 | 7 | 4 | 0 | 10 | 1 | 4 | 6 | 5 |
| | **Mean** | 0.112 | 0.248 | 0.128 | 0.122 | 0.183 | 0.074 | 0 | 0.023 | 0.004 | 0.248 | 0.170 | 0.025 |
| | Max | 0.657 | 0.429 | 0.486 | 0.231 | 0.442 | 0.274 | 0 | 0.162 | 0.029 | 0.653 | 0.530 | 0.173 |
| | Min | 0.126 | 0.277 | 0.109 | 0.202 | 0.193 | 0.244 | 0 | 0.162 | 0.029 | 0.170 | 0.298 | 0.173 |
| **Computer** | **Duration** | 4.36 | 3.75 | 3.33 | 2.11 | 4.44 | 1.21 | 0 | 0.71 | 1.80 | 5.48 | 1.96 | 0.01 |
| | Max | 2.31 | 1.38 | 1.50 | 0.72 | 1.39 | 0.80 | 0 | 0.71 | 1.17 | 1.48 | 1.05 | 0.01 |
| | Min | 0.08 | 0.29 | 0.44 | 0.29 | 0.73 | 0.40 | 0 | 0.71 | 0.01 | 0.64 | 0.02 | 0.01 |
| | **Changes** | 3 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 0 |
| | Max | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| | Min | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| | **Mean** | 0.037 | 0.016 | 0.013 | 0.000 | 0.000 | 0.015 | 0 | 0 | 0 | 0.026 | 0.019 | 0 |
| | Max | 0.151 | 0.110 | 0.094 | 0.000 | 0.000 | 0.106 | 0 | 0 | 0 | 0.094 | 0.134 | 0 |
| | Min | 0.109 | 0.110 | 0.094 | 0.000 | 0.000 | 0.106 | 0 | 0 | 0 | 0.088 | 0.134 | 0 |
| **Window** | **Duration** | 0 | 0 | 0 | 4.68 | 3.32 | 2.05 | 0 | 0 | 3.62 | 13.42 | 13.84 | 0 |
| | Max | 0 | 0 | 0 | 2.41 | 3.21 | 2.05 | 0 | 0 | 3.59 | 3.86 | 5.09 | 0 |
| | Min | 0 | 0 | 0 | 2.27 | 0.11 | 2.05 | 0 | 0 | 0.03 | 1.78 | 4.15 | 0 |
| | **Changes** | 0 | 0 | 0 | 2 | 2 | 1 | 0 | 0 | 2 | 5 | 3 | 0 |
| | Max | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 |
| | Min | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 |
| | **Mean** | 0 | 0 | 0 | 0.668 | 0.475 | 0.293 | 0 | 0 | 0.517 | 1.917 | 1.978 | 0 |
| | Max | 0 | 0 | 0 | 2.409 | 3.208 | 2.050 | 0 | 0 | 3.594 | 3.860 | 5.086 | 0 |
| | Min | 0 | 0 | 0 | 2.267 | 0.115 | 2.050 | 0 | 0 | 0.027 | 1.782 | 4.154 | 0 |

Table 4: The summary of statistical results of room temperature and humidity in office environment. The data labelled $D4$, $D5$ and $D6$ as shown in Table 3.2 are used.

| | | User #1 | | | | | User #2 | | | | | User #4 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Weeks | Mon | Teu | Wed | Thu | Fri | Mon | Teu | Wed | Thu | Fri | Mon | Teu | Wed | Thu | Fri |
| Indoor temperature (°C) | 1 | 18.58 | 18.87 | 20.37 | 20.87 | 19.84 | 21.56 | 23.62 | 24.25 | 22.60 | 23.18 | 21.06 | 23.70 | 23.78 | 23.28 | 22.92 |
| | 2 | 16.50 | 19.48 | 19.65 | 20.94 | 19.77 | 20.05 | 21.98 | 23.57 | 23.24 | 21.70 | 20.22 | 22.38 | 22.36 | 22.02 | 21.60 |
| | 3 | 17.87 | 19.47 | 21.07 | 20.65 | 20.81 | 20.38 | 23.35 | 24.21 | 24.26 | 24.48 | 21.47 | 23.37 | 24.17 | 23.20 | 23.24 |
| | 4 | 18.15 | 20.17 | 21.04 | 18.78 | 19.79 | 21.33 | 24.53 | 24.37 | 24.28 | 24.43 | 19.88 | 21.45 | 21.69 | 22.35 | 22.96 |
| | 5 | 19.58 | 20.57 | 20.88 | 21.43 | 20.77 | 22.44 | 24.27 | 23.99 | 24.53 | 24.98 | 19.57 | 24.13 | 22.74 | 24.55 | 25.17 |
| | 6 | 20.13 | 20.06 | 20.92 | 21.87 | 20.41 | 22.59 | 24.12 | 24.16 | 23.65 | 23.37 | 21.11 | 22.85 | 22.27 | 21.94 | 21.26 |
| | 7 | 21.23 | 21.86 | 21.08 | 21.41 | 23.51 | 23.21 | 24.08 | 23.11 | 25.43 | 25.40 | 20.38 | 20.86 | 18.78 | 24.80 | 24.87 |
| | 8 | 21.23 | 22.04 | 23.47 | 23.96 | 24.21 | 23.21 | 23.81 | 25.61 | 25.80 | 25.71 | 20.38 | 22.60 | 25.23 | 24.15 | 24.87 |
| | 9 | 16.87 | 20.31 | 22.10 | 21.45 | 22.38 | 21.82 | 22.60 | 23.51 | 22.72 | 24.72 | 18.05 | 23.22 | 20.91 | 21.86 | 23.68 |
| | 10 | 21.51 | 23.56 | 22.65 | 21.80 | 22.05 | 24.39 | 24.69 | 25.34 | 24.61 | 26.06 | 24.45 | 25.10 | 24.94 | 25.27 | 25.38 |
| | 11 | 19.74 | 22.78 | 24.35 | 23.25 | 23.48 | 23.55 | 23.66 | 23.88 | 23.11 | 23.21 | 22.91 | 24.35 | 23.93 | 23.48 | 24.33 |
| | 12 | 20.34 | 22.56 | 22.45 | 22.49 | 22.80 | 21.49 | 24.33 | 24.18 | 23.52 | 23.55 | 22.51 | 24.74 | 23.70 | 22.70 | 22.84 |
| Indoor humidity ($\%RH$) | 1 | 43.10 | 42.67 | 42.86 | 40.69 | 35.62 | 38.27 | 35.17 | 37.70 | 37.44 | 31.21 | 39.73 | 37.69 | 37.64 | 33.75 | 31.67 |
| | 2 | 33.57 | 30.40 | 37.36 | 37.17 | 36.95 | 30.27 | 28.41 | 36.05 | 34.24 | 33.77 | 29.63 | 27.85 | 35.10 | 33.76 | 34.08 |
| | 3 | 36.95 | 36.43 | 40.55 | 38.43 | 33.43 | 34.86 | 32.60 | 37.16 | 33.33 | 29.23 | 33.23 | 33.60 | 36.69 | 33.08 | 29.60 |
| | 4 | 33.94 | 29.92 | 28.14 | 27.64 | 27.54 | 29.10 | 26.18 | 24.66 | 22.80 | 23.75 | 30.61 | 27.10 | 22.53 | 23.69 | 24.22 |
| | 5 | 33.19 | 29.98 | 26.86 | 27.02 | 26.92 | 28.24 | 27.26 | 23.49 | 22.58 | 24.32 | 29.07 | 26.80 | 21.82 | 22.86 | 23.24 |
| | 6 | 31.24 | 30.01 | 31.85 | 33.58 | 35.35 | 26.88 | 26.56 | 27.35 | 29.34 | 31.08 | 28.29 | 28.76 | 31.48 | 33.97 | 37.61 |
| | 7 | 38.27 | 34.63 | 36.18 | 40.62 | 39.32 | 33.39 | 31.04 | 31.22 | 37.41 | 34.19 | 36.67 | 38.78 | 43.79 | 39.75 | 34.93 |
| | 8 | 38.27 | 40.82 | 39.46 | 35.70 | 34.24 | 33.39 | 37.16 | 35.23 | 31.66 | 29.06 | 36.67 | 39.67 | 34.10 | 32.57 | 31.41 |
| | 9 | 31.60 | 28.40 | 30.54 | 29.15 | 32.32 | 27.68 | 26.84 | 31.02 | 28.00 | 32.02 | 31.34 | 27.01 | 33.97 | 29.91 | 32.55 |
| | 10 | 34.47 | 33.80 | 34.24 | 33.06 | 32.95 | 32.12 | 32.90 | 32.34 | 30.44 | 33.11 | 33.48 | 31.66 | 30.99 | 30.00 | 30.59 |
| | 11 | 32.21 | 32.82 | 31.84 | 33.34 | 35.65 | 30.34 | 31.94 | 31.33 | 32.42 | 35.82 | 28.66 | 30.61 | 30.72 | 33.55 | 36.72 |
| | 12 | 34.01 | 31.36 | 28.97 | 28.94 | 31.35 | 34.01 | 29.90 | 29.50 | 30.11 | 33.48 | 33.13 | 28.72 | 29.33 | 30.55 | 32.89 |
| Outdoor temperature (°C) | 1 | 8.24 | 8.76 | 10.71 | 8.32 | 3.35 | 8.24 | 8.76 | 10.71 | 8.32 | 3.35 | 8.24 | 8.76 | 10.71 | 8.32 | 3.35 |
| | 2 | 0.15 | 0.82 | 7.63 | 6.41 | 6.74 | 0.15 | 0.82 | 7.63 | 6.41 | 6.74 | 0.15 | 0.82 | 7.63 | 6.41 | 6.74 |
| | 3 | 5.31 | 6.17 | 9.84 | 6.13 | 4.77 | 5.31 | 6.17 | 9.84 | 6.13 | 4.77 | 5.31 | 6.17 | 9.84 | 6.13 | 4.77 |
| | 4 | 2.18 | 1.19 | 0.32 | 0.67 | -0.13 | 2.18 | 1.19 | 0.32 | 0.67 | -0.13 | 2.18 | 1.19 | 0.32 | 0.67 | -0.13 |
| | 5 | 4.38 | 1.20 | 0.16 | 2.22 | 1.17 | 4.38 | 1.20 | 0.16 | 2.22 | 1.17 | 4.38 | 1.20 | 0.16 | 2.22 | 1.17 |
| | 6 | 5.79 | 6.78 | 8.20 | 8.34 | 9.09 | 5.79 | 6.78 | 8.20 | 8.34 | 9.09 | 5.79 | 6.78 | 8.20 | 8.34 | 9.09 |
| | 7 | 10.21 | 9.95 | 8.53 | 13.07 | 10.78 | 10.21 | 9.95 | 8.53 | 13.07 | 10.78 | 10.21 | 9.95 | 8.53 | 13.07 | 10.78 |
| | 8 | 10.21 | 12.27 | 11.01 | 9.66 | 8.89 | 10.21 | 12.27 | 11.01 | 9.66 | 8.89 | 10.21 | 12.27 | 11.01 | 9.66 | 8.89 |
| | 9 | 5.11 | 5.99 | 7.68 | 7.87 | 11.12 | 5.11 | 5.99 | 7.68 | 7.87 | 11.12 | 5.11 | 5.99 | 7.68 | 7.87 | 11.12 |
| | 10 | 8.11 | 8.93 | 8.02 | 6.47 | 9.29 | 8.11 | 8.93 | 8.02 | 6.47 | 9.29 | 8.11 | 8.93 | 8.02 | 6.47 | 9.29 |
| | 11 | 9.94 | 10.48 | 9.98 | 8.04 | 9.75 | 9.94 | 10.48 | 9.98 | 8.04 | 9.75 | 9.94 | 10.48 | 9.98 | 8.04 | 9.75 |
| | 12 | 10.27 | 12.41 | 13.54 | 13.04 | 11.18 | 10.27 | 12.41 | 13.54 | 13.04 | 11.18 | 10.27 | 12.41 | 13.54 | 13.04 | 11.18 |

# Appendix E - Similarity Measure Results of Users' Behaviour

Detailed experimental results of similarity measures for users' behaviour are presented in this appendix.

Table 1: The similarity measure results of User #2 activity in an office environment for periods 5 days and 5 weeks. The similarity measures are used Hamming Distance technique.

| Data | 43200 samples per day for a period of 5 days | | | | | |
|---|---|---|---|---|---|---|
| | | | Office occupancy | | | |
| Day | 1 | 2 | 3 | 4 | 5 | Total |
| 1 | 0 | 8940 | 16601 | 6601 | 14504 | 46646 |
| 2 | 8940 | 0 | 9300 | 2106 | 17795 | 38141 |
| 3 | 16601 | 9300 | 0 | 1399 | 11978 | 39278 |
| 4 | **6601** | **2106** | **1399** | **0** | **17924** | **28031** |
| 5 | 14504 | 17795 | 11978 | 17924 | 0 | 62202 |
| | | | Light activity | | | |
| | 1 | 2 | 3 | 4 | 5 | Total |
| 1 | 0 | 8035 | 21911 | 5920 | 14260 | 50125 |
| 2 | 8035 | 0 | 13876 | 0 | 13770 | 35681 |
| 3 | 21911 | 13876 | 0 | 0 | 2635 | 38421 |
| 4 | **5920** | **0** | **0** | **0** | **13968** | **19888** |
| 5 | 14260 | 13770 | 2635 | 13968 | 0 | 44633 |
| | | | Computer activity | | | |
| | 1 | 2 | 3 | 4 | 5 | Total |
| 1 | **0** | **4886** | **5326** | **4823** | **5245** | **20280** |
| 2 | 4886 | 0 | 6929 | 6737 | 8334 | 26886 |
| 3 | 5326 | 6929 | 0 | 4350 | 5498 | 22103 |
| 4 | 4823 | 6737 | 4350 | 0 | 7602 | 23511 |
| 5 | 5245 | 8334 | 5498 | 7602 | 0 | 26679 |
| Data | 216000 samples per week for a period of 5 weeks | | | | | |
| | | | Office occupancy | | | |
| Week | 1 | 2 | 3 | 4 | 5 | Total |
| 1 | **0** | **41094** | **42293** | **38042** | **58537** | **179966** |
| 2 | 41094 | 0 | 51125 | 60687 | 67915 | 220821 |
| 3 | 42293 | 51125 | 0 | 53101 | 57040 | 203559 |
| 4 | 38042 | 60687 | 53101 | 0 | 34551 | 186381 |
| 5 | 58537 | 67915 | 57040 | 34551 | 0 | 218043 |
| | | | Light activity | | | |
| | 1 | 2 | 3 | 4 | 5 | Total |
| 1 | 0 | 30761 | 41014 | 27101 | 25979 | 124855 |
| 2 | 30761 | 0 | 38553 | 32947 | 29394 | 131655 |
| 3 | 41014 | 38553 | 0 | 29394 | 30307 | 139268 |
| 4 | 27101 | 32947 | 29394 | 0 | 24106 | 113548 |
| 5 | **25979** | **29394** | **30307** | **24106** | **0** | **109786** |
| | | | Computer activity | | | |
| | 1 | 2 | 3 | 4 | 5 | Total |
| 1 | 0 | 25661 | 24516 | 23638 | 26274 | 100089 |
| 2 | **25661** | **0** | **21469** | **23879** | **25016** | **96025** |
| 3 | 24516 | 21469 | 0 | 29250 | 29669 | 104904 |
| 4 | 23638 | 23879 | 29250 | 0 | 21518 | 98285 |
| 5 | 26274 | 25016 | 29669 | 21518 | 0 | 102477 |

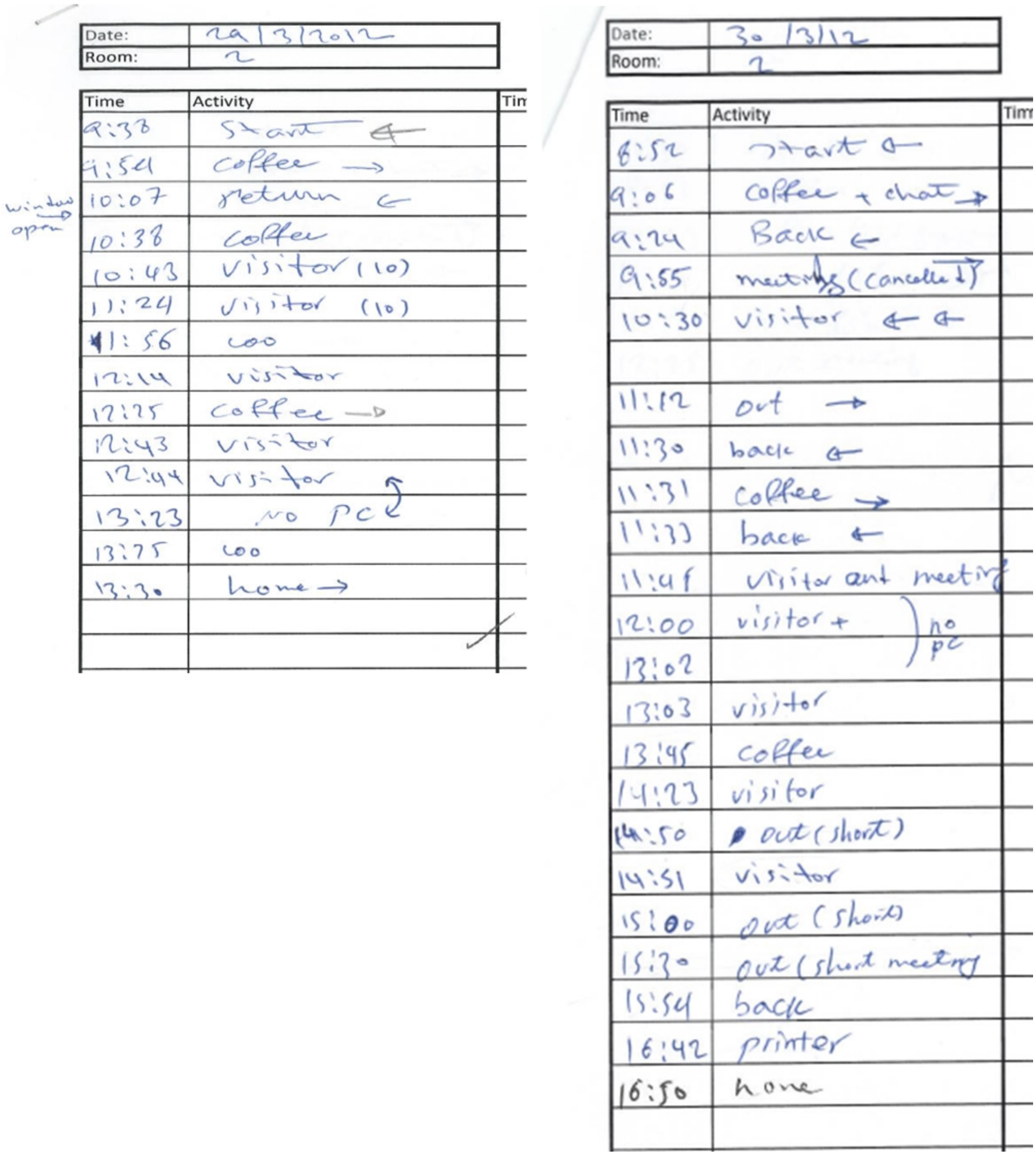
# Appendix F - Validation of Activity Recognition

Detailed annotation activity and fuzzy rules of single-user's activity recognition in an office environment are presented in this appendix.

In this research, data $D7$ from User #3 was used to validate this activity recognition model. The validation compares the workers activity, detected by the recognition model, with the annotated data. The percentage obtained from comparing a recognition signal with a sensory signal was used to compute the accuracy of the model.

An annotation method was used to record workers' activity during a working day [168, 169], for comparison between the WSN data collection and the output from the activity recognition model. In order to compare the sensory data and the manually recorded data, the sensors data was collected continuously using the office intelligence system and the specific ADW of users in an office environment was annotated. For example, Figures 1 and Figure 2 show specific activities that were manually recorded for Users #2 and #3, respectively.

In this research, activity recognition models were developed to recognise and classify the ADL of an office worker into six categories of user's activity (e.g.



| Date: | 29/3/2012 |
|---|---|
| Room: | 2 |

| Time | Activity | Tim |
|---|---|---|
| 9:38 | start ← | |
| 9:54 | coffee → | |
| 10:07 | return ← | |
| 10:38 | coffee | |
| 10:43 | visitor (10) | |
| 11:24 | visitor (10) | |
| 11:56 | loo | |
| 12:14 | visitor | |
| 12:25 | coffee → | |
| 12:43 | visitor | |
| 12:44 | visitor | |
| 13:23 | no PC | |
| 13:25 | loo | |
| 13:30 | home → | |
| | | |
| | | |

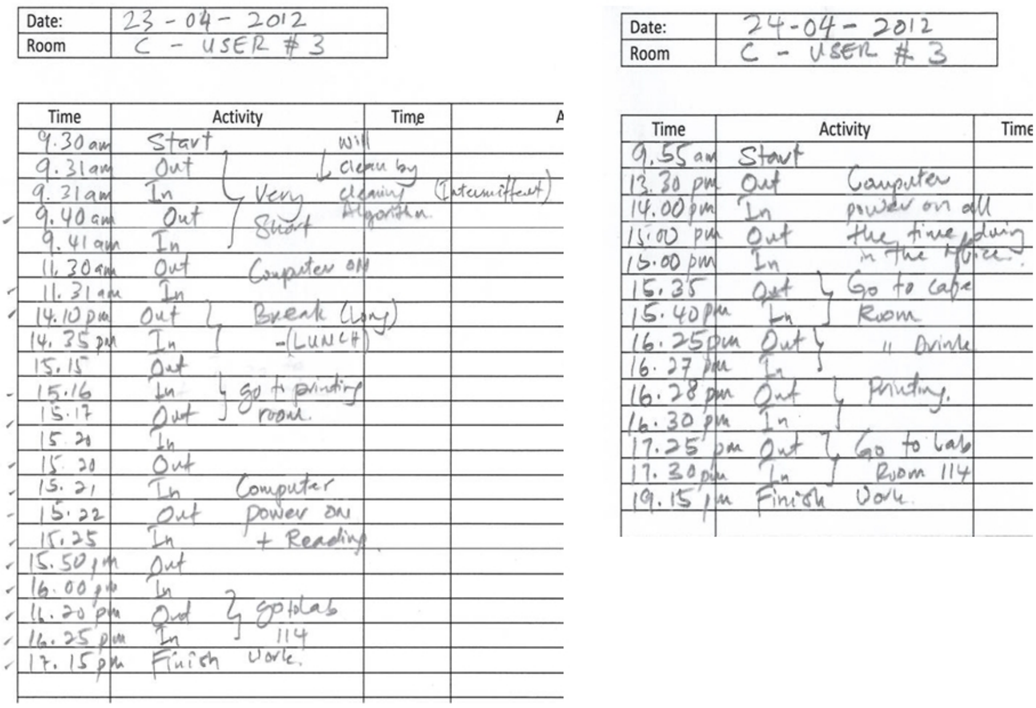
window open

| Date: | 30/3/12 |
|---|---|
| Room: | 2 |

| Time | Activity | Tim |
|---|---|---|
| 8:52 | start ← | |
| 9:06 | coffee + chat → | |
| 9:24 | Back ← | |
| 9:55 | meeting (cancelled) | |
| 10:30 | visitor ← ← | |
| | | |
| 11:12 | out → | |
| 11:30 | back ← | |
| 11:31 | coffee → | |
| 11:33 | back ← | |
| 11:45 | visitor and meeting | |
| 12:00 | visitor + no pc | |
| 13:02 | | |
| 13:03 | visitor | |
| 13:45 | coffee | |
| 14:23 | visitor | |
| 14:50 | out (short) | |
| 14:51 | visitor | |
| 15:00 | out (short) | |
| 15:30 | out (short meeting | |
| 15:54 | back | |
| 16:42 | printer | |
| 16:50 | home | |
| | | |

Figure 1: The record forms of user's activity annotation in an office environment for User #2.

at home, out of office for duties, lunch, short, not using computer, and using computer). Example patterns of activity recognition are shown in Figure 3. In the university's office environment, accurate activity recognition was challenging, because the academic office workers (as the observed users of this research) did



| Date: | 23 - 04 - 2012 |
|---|---|
| Room | C - USER #3 |

| Time | Activity | Time | A |
|---|---|---|---|
| 9.30 am | Start will | | |
| 9.31 am | Out clean by | | |
| 9.31 am | In Very cleaning (Intermittent) | | |
| 9.40 am | Out Short Algorithm. | | |
| 9.41 am | In | | |
| 11.30 am | Out Computer off | | |
| 11.31 am | In | | |
| 14.10 pm | Out Break (Long) | | |
| 14.35 pm | In -(LUNCH) | | |
| 15.15 | Out | | |
| 15.16 | In Go to printing | | |
| 15.17 | Out room. | | |
| 15.20 | In | | |
| 15.20 | Out | | |
| 15.21 | In Computer | | |
| 15.22 | Out power on | | |
| 15.25 | In + Reading | | |
| 15.50 pm | Out | | |
| 16.00 pm | In | | |
| 16.20 pm | Out Go to lab | | |
| 16.25 pm | In 114 | | |
| 17.15 pm | Finish Work. | | |

| Date: | 24-04-2012 |
|---|---|
| Room | C - USER #3 |

| Time | Activity | Time |
|---|---|---|
| 9.55 am | Start | |
| 13.30 pm | Out Computer | |
| 14.00 pm | In power on all | |
| 15.00 pm | Out the time during | |
| 15.00 pm | In in the office. | |
| 15.35 | Out Go to cafe | |
| 15.40 pm | In Room | |
| 16.25 pm | Out " Drink | |
| 16.27 pm | In | |
| 16.28 pm | Out Printing. | |
| 16.30 pm | In | |
| 17.25 pm | Out Go to lab | |
| 17.30 pm | In Room 114 | |
| 19.15 pm | Finish Work. | |

Figure 2: The record forms of user's activity annotation in an office environment for User #3.

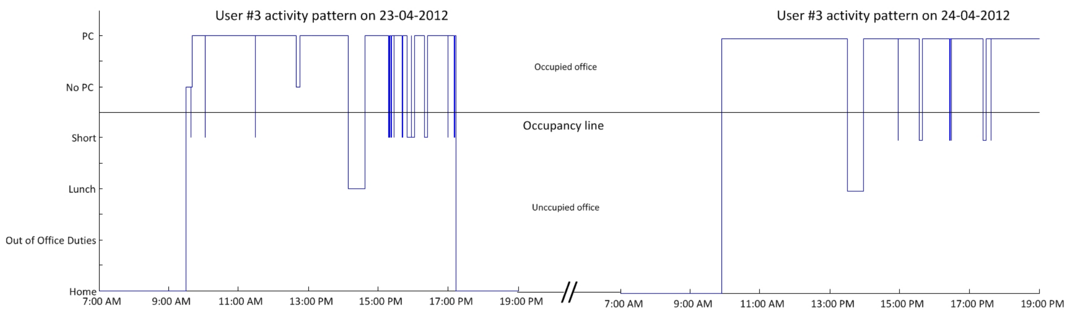



Figure 3: The samples of activity patterns, generated by activity recognition of User #3 for two days (23-04-2012 to 24-04-2012).

not use the office all of the time (i.e., they were lecturing, in meetings, temporarily absent during semester breaks etc.).

In Figure 4, the comparisons between collected sensory signals and recognition output signals were used to compute the accuracy of the activity recognition. The

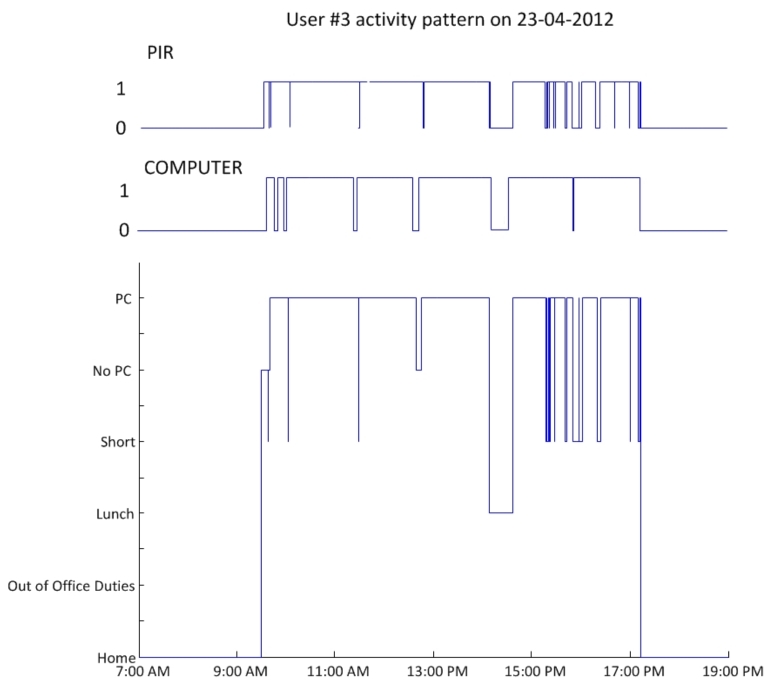



Figure 4: The comparision between sensory signals and recognition out signal.

accuracy of activity recognition, using an event-driven model and IF-THEN rules for real data based on User #3s activities on 23-04-2012, obtained 91 % (i.e., between Computer's signal and user's activity signal of using PC and not using PC). Meanwhile, the accuracy between PIR's signal and user's activity of out of the office, such as at home, out of office duties, lunch, and short breaks, was 90.81 % (i.e., between PIR and user's activity signals). Further examination showed that overlapping occurred between different user's activities, which influenced the accuracy of the activity recognition. For example, if the computer was still turned on, but the user had left the office, the activity recognition gave priority to leaves the office as being the selected activity. Therefore, the comparison between sensors signals and recognition signals, slightly affected the percentage of recognition accuracy.

# References

[1] L. Narayanan, S. Menon, and P. E. Spector, "Stress in the workplace: a comparison of gender and occupations", *Journal of Organizational Behavior*, vol. 20, pp. 63–73, 1999. 1

[2] A. I. Dounis and C. Caraiscos, "Advanced control systems engineering for energy and comfort management in a building environmental - a review", *Renewable and Sustainable Energy Reviews*, vol. 13, no. 6-7, pp. 1246–1261, 2009. 2, 13

[3] Mohd Fauzi Othman and Khairunnisa Shazali, "Wireless sensor network applications: A study in environment monitoring system", *Procedia Engineering*, vol. 41, pp. 1204–1210, 2012. 2, 26

[4] Kelly R. Cowan and Tugrul U. Daim, "Review of technology acquisition and adoption research in the energy sector", *Technology in Society*, vol. 33, no. 3-4, pp. 183–199, 2011. 2

[5] Liming Chen, J. Hoey, C. D. Nugent, D. J. Cook, and Zhiwen Yu, "Sensor-based activity recognition", *IEEE Transactions on Systems and Cybernetics, Part C: Applications and Reviews*, vol. 42, no. 6, pp. 790–808, 2012. 2, 16, 25

[6] Andreas Berl and Hermann de Meer, "An energy consumption model for virtualized office environments", *Future Generation Computer Systems*, vol. 27, no. 8, pp. 1047–1055, 10 2011. 9

[7] Koen Steemers and Shweta Manchanda, "Energy efficient design and occupant well-being: Case studies in the UK and India", *Building and Environment*, vol. 45, no. 2, pp. 270–278, 2 2010. 9, 23

[8] Li-Chien Huang, Hong-Chan Chang, Cheng-Chung Chen, and Cheng-Chien Kuo, “A Zigbee-based monitoring and protection system for building electrical safety”, *Energy and Buildings*, vol. 43, no. 6, pp. 1418–1426, 6 2011. 9, 23

[9] Barry P. Haynes, “The impact of office comfort on productivity”, *Journal of facilities management*, vol. 6, no. 1, pp. 51–37, 2008. 9, 14

[10] Department of Energy and Climate Change of UK, “Energy efficiency deployment office”, 2012. 10, 11

[11] Ayman Khalid Elsadig, “Energy efficiency in commercial buildings”, *A dissertation: Master of Science Sustainable Engineering: Energy Systems and the Environment, University of Strathclyde, U.K*, 2005. 10

[12] H. Hagras, I. Packharn, Y. Vanderstockt, N. McNulty, A. Vadher, and F. Doctor, “An intelligent agent based approach for energy management in commercial buildings”, in *Proc. of the IEEE World Congress on Computational Intelligence*, 2008, pp. 156–162. 10, 31

[13] William Chung and Y. V. Hui, “A study of energy efficiency of private office buildings in Hong Kong”, *Energy and Buildings*, vol. 41, no. 6, pp. 696–701, 2009. 10

[14] S. Ari, P. Wilcoxen, H.E. Khalifa, J.F. Dannenhoffer, and C. Isik, “A practical approach to individual thermal comfort and energy optimization problem”, in *Annual Meeting of the North American on Fuzzy Information Processing Society (NAFIPS)*, may 2008, pp. 1–6. 11, 13, 29

[15] D. Jenkins, Y. Liu, and A. D. Peacock, “Climatic and internal factors affecting future UK office heating and cooling energy consumptions”, *Energy and Buildings*, vol. 40, no. 5, pp. 874–81, 2008. 11

[16] Philomena M. Bluyssen, Myriam Aries, and Paula van Dommelen, “Comfort of workers in office buildings: The European HOPE project”, *Building and Environment*, vol. 46, no. 1, pp. 280–288, 1 2011. 11

[17] MERLSense, “Two years in a tar file”, *CCOHS [online] Available: http://www.ccohs.ca/oshanswers/phys_agents/thermal_comfort.html*. 12

[18] CASAS Smart Home Project, "Smart home datasets", *Washington State University [online] Available: http://ailab.wsu.edu/casas/datasets.html*. 12

[19] John Fischer, Kirk Mescher, Ben Elkin, Stephen M. McCune, and Jack Gresham, "High-performance schools: High marks for energy efficiency, humidity control, indoor air quality first cost", *ASHRAE Journal*, vol. 49, no. 5, pp. 30–46, 2007. 12, 13

[20] Yong Razidah Rashid, Mohd Sabere Sulaiman, Azlina Aziz, Hilmilia Selamat, Abdul Halim Mat Yani, and Mohd Zin Kandar, "Greening government's office buildings: PWD Malaysia experiences", *Procedia Engineering*, vol. 21, pp. 1056–1060, 2011. 12

[21] Stuart Barlow and Dusan Fiala, "Occupant comfort in UK officeshow adaptive comfort theories might influence future low energy office refurbishment strategies", *Energy and Buildings*, vol. 39, no. 7, pp. 837–846, 7 2007. 12

[22] Chartered Institution of Building Services Engineers, "Energy use in offices", *Energy Consumption Guide 19*, 2000. 12

[23] Commercial Buildings Energy Consumption Survey, "Overview of commercial buildings - Full report", *Commercial Buildings Energy Consumption Survey, United State*, 2003. 12

[24] Tugce Kazanasmaz, Murat Gunaydin, and Selcen Binol, "Artificial neural networks to predict daylight illuminance in office buildings", *Building and Environment*, vol. 44, no. 8, pp. 1751–1757, 2009. 12, 30

[25] S. L. Wong, K. K. W. Wan, and T. N. T. Lam, "Artificial neural networks for energy analysis of office buildings with daylighting", *Applied Energy*, vol. 87, no. 2, pp. 551–7, 02 2010. 12, 30

[26] B. B. Ekici and U. T. Aksoy, "Prediction of building energy consumption by using artificial neural networks", *Adv. Eng. Software*, vol. 40, pp. 356–362, 2009. 13

[27] Nyuk Hien Wong and Shan Shan Khoo, "Thermal comfort in classrooms in the tropics", *Energy and Buildings*, vol. 35, no. 4, pp. 337–351, 2003. 13

[28] N. Gaitani, C. Lehmann, M. Santamouris, G. Mihalakakou, and P. Patargias, "Using principal component and cluster analysis in the heating evaluation of the school building sector", *Applied Energy*, vol. 87, no. 6, pp. 2079–2086, 2010. 13

[29] Heschong Mahone Group, "Windows and offices: A study of office worker performance and the indoor environment", *California Energy Commission's Public Interest Energy Research (CEC PIER)*, 2003. 13

[30] Olli Seppanen, William J. Fisk, and Q.H. Lei, "Effect of temperature on task performance in office environment", *Lawrence Berkeley National Laboratory*, 2006. 13

[31] M. B. C. Aries, J. A. Veitch, and G. R. Newsham, "Physical and psychological discomfort in the office environment", *Light, Performance and Quality of Life – Light and Health Research Foundation (SOLG) Symposium*, 2007. 13

[32] Michael O'Neill, "A model of environmental control and effective work", *Knoll Workplace Research*, vol. Knoll, Inc., 2008. 14

[33] P. Rashidi, D.J. Cook, L.B. Holder, and M. Schmitter-Edgecombe, "Discovering activities to recognize and track in a smart environment", *Knowledge and Data Engineering, IEEE Transactions on*, vol. 23, no. 4, pp. 527 –539, april 2011. 14, 16, 18, 28, 63

[34] Jing Liu, Runming Yao, and Rachel McCloy, "A method to weight three categories of adaptive thermal comfort", *Energy and Buildings*, vol. 47, pp. 312–320, 4 2012. 14, 23

[35] Runming Yao, Jing Liu, and Baizhan Li, "Occupants' adaptive responses and perception of thermal environment in naturally conditioned university classrooms", *Applied Energy*, vol. 87, no. 3, pp. 1015–22, 03 2010. 14

[36] P.O. Fanger, *Thermal comfort: analysis and applications in environmental engineering*, McGraw-Hill, 1970. 14

[37] Shengxian Wei, Ming Li, Wenxian Lin, and Yanlin Sun, "Parametric studies and evaluations of indoor thermal environment in wet season using a field survey and pmv/ppd method", *Energy and Buildings*, vol. 42, no. 6, pp. 799–806, 6 2010. 14

[38] Guodong Ye, Changzhi Yang, Youming Chen, and Yuguo Li, "A new approach for measuring predicted mean vote (pmv) and standard effective temperature (set)", *Building and Environment*, vol. 38, no. 1, pp. 33–44, 1 2003. 14

[39] Canadian centre for Occupantional health and Safety, "Thermal comfort for office work", *Mitsubishi Electric Research Laboratories [online] Available: http://www.merl.com/wmd/*. 14, 93

[40] ASHRAE Press, *Air Conditioning System Design Manual (Second Edition)*, pp. 1–7, Butterworth-Heinemann, Burlington, 2007. 14

[41] Hugo S. L. C. Hens, "Thermal comfort in office buildings: Two case studies commented", *Building and Environment*, vol. 44, no. 7, pp. 1399–1408, 7 2009. 14

[42] Yaning Wang and Zhaofeng Wang, "Design of intelligent residential lighting control system based on zigbee wireless sensor network and fuzzy controller", in *Machine Vision and Human-Machine Interface (MVHI), 2010 International Conference on*, 2010, pp. 561–564. 15, 29, 40

[43] Jiang He, Akira Hoyano, and Takashi Asawa, "A numerical simulation tool for predicting the impact of outdoor thermal environment on building energy performance", *Applied Energy*, vol. 86, no. 9, pp. 1596–1605, 9 2009. 15

[44] Ashfaque Ahmed Chowdhury, M. G. Rasul, and M. M. K. Khan, "Thermal-comfort analysis and simulation for various low-energy cooling-technologies applied to an office building in a subtropical climate", *Applied Energy*, vol. 85, no. 6, pp. 449–462, 2008. 15

[45] Zhang Lin, T. T. Chow, K. F. Fong, Qiuwang Wang, and Ying Li, "Comparison of performances of displacement and mixing ventilations. Part I:

thermal comfort", *International Journal of Refrigeration*, vol. 28, no. 2, pp. 276–287, 3 2005. 15, 26

[46] Johnny Wong, Heng Li, and Jenkin Lai, "Evaluating the system intelligence of the intelligent building systems: Part 1: Development of key intelligent indicators and conceptual analytical framework", *Automation in Construction*, vol. 17, no. 3, pp. 284–302, 3 2008. 15

[47] Johnny K. W. Wong and Heng Li, "Application of the analytic hierarchy process (AHP) in multi-criteria analysis of the selection of intelligent building systems", *Building and Environment*, vol. 43, no. 1, pp. 108–125, 1 2008. 15

[48] Johnny Wong and Heng Li, "Development of a conceptual model for the selection of intelligent building systems", *Building and Environment*, vol. 41, no. 8, pp. 1106–1123, 8 2006. 15

[49] Haris Doukas, Christos Nychtis, and John Psarras, "Assessing energy-saving measures in buildings through an intelligent decision support model", *Building and Environment*, vol. 44, no. 2, pp. 290–298, 2009. 15, 16

[50] N. K. Suryadevara, A. Gaddam, R. K. Rayudu, and S. C. Mukhopadhyay, "Wireless sensors network based safe home to care elderly people: Behaviour detection", *Sensors and Actuators A: Physical*, vol. 186, pp. 277–283, 10 2012. 16, 17

[51] Long Xiao, Bo Cheng, Bo Yang, Rong Du, Wenbin Yu, and Xinping Guan, "A context-aware entrance guard in smart home: An event-driven application based on the human motion and face recognition", in *Automation, Robotics and Applications (ICARA), 2011 5th International Conference on*, 2011, pp. 184–189. 16

[52] Chao Chen, B. Das, and D. J. Cook, "A data mining framework for activity recognition in smart environments", in *Intelligent Environments (IE), 2010 Sixth International Conference on*, 2010, pp. 80–83. 16

[53] E. Nazerfard, P. Rashidi, and D.J. Cook, "Discovering temporal features and relations of activity patterns", in *Data Mining Workshops (ICDMW),*

*2010 IEEE International Conference on*, dec. 2010, pp. 1069 –1075. 16, 21, 28, 63

[54] Hamid Medjahed, Dan Istrate, Jerome Boudy, and Bernadette Dorizzi, "Human activities of daily living recognition using fuzzy logic for elderly home monitoring", in *Proc. of the International Conference on Fuzzy Systems*, Republic of Korea, August 2009, pp. 2001–2006. 16, 17, 29

[55] Hermie J. Hermens and Miriam M. R. Vollenbroek-Hutten, "Towards remote monitoring and remotely supervised training", *Journal of Electromyography and Kinesiology*, vol. 18, no. 6, pp. 908–919, 12 2008. 16, 17

[56] Vincent Rialle, Jean-Baptiste Lamy, Norbert Noury, and Lionel Bajolle, "Telemonitoring of patients at home: a software agent approach", *Computer methods and programs in biomedicine*, vol. 72, no. 3, pp. 257–268, 11 2003. 16

[57] Yuanlong Liu, "Wireless remote patient monitoring on general hospital wards", *Ph.D.- Bournemouth University, Poole*, 2010. 16

[58] Venet Osmani, Sasitharan Balasubramaniam, and Dmitri Botvich, "Human activity recognition in pervasive health-care: Supporting efficient remote collaboration", *Journal of Network and Computer Applications*, vol. 31, no. 4, pp. 628–655, 11 2008. 16, 17, 18

[59] Saifullizam Puteh, Caroline Langensiepen, and Ahmad Lotfi, "Similarity pattern mining in intelligent office environments", *in Workshop Proceedings of the 7th International Conference on Intelligent Environments*, vol. 10, pp. 562–573, 2011. 16

[60] S. Puteh, C. Langensiepen, and A. Lotfi, "Fuzzy ambient intelligence for intelligent office environments", in *Fuzzy Systems (FUZZ-IEEE), 2012 IEEE International Conference on*, 2012, pp. 1–6. 16, 63

[61] Saifullizam Puteh, Caroline Langensiepen, and Ahmad Lotfi, "University office simulator for energy and comfort optimisation", *the 25th European Conference on Modelling and Simulation (ECMS 2011)*, pp. 430–435, 2011. 16

[62] Eunju Kim, Sumi Helal, and D. Cook, "Human activity recognition and pattern discovery", *IEEE Pervasive Computing*, vol. 9, no. 1, pp. 48–53, 01 2010. 17, 63

[63] Vincent Tabak and Bauke de Vries, "Methods for the prediction of intermediate activities by office occupants", *Building and Environment*, vol. 45, no. 6, pp. 1366–1372, 2010. 17

[64] Xin Hong and C. D. Nugent, "Partitioning time series sensor data for activity recognition", in *Information Technology and Applications in Biomedicine, 2009. ITAB 2009. 9th International Conference on*, 2009, pp. 1–4. 17

[65] Saisakul Chernbumroong, Shuang Cang, Anthony Atkins, and Hongnian Yu, "Elderly activities recognition and classification for applications in assisted living", *Expert Systems with Applications*, vol. 40, no. 5, pp. 1662–1674, 4 2013. 17, 23

[66] A. H. Shabani, D. A. Clausi, and J. S. Zelek, "Evaluation of local spatio-temporal salient feature detectors for human action recognition", in *Computer and Robot Vision (CRV), 2012 Ninth Conference on*, 2012, pp. 468–475. 17

[67] Fco Javier Ordez, Jos Antonio Iglesias, Paula de Toledo, Agapito Ledezma, and Araceli Sanchis, "Online activity recognition using evolving classifiers", *Expert Systems with Applications*. 17

[68] D. Naranjo-Hernandez, L. M. Roa, J. Reina-Tosina, and M. A. Estudillo-Valderrama, "SoM: A smart sensor for human activity monitoring and assisted healthy ageing", *Biomedical Engineering, IEEE Transactions on*, vol. 59, no. 11, pp. 3177–3184, 2012. 18, 23, 24

[69] R. R. Fletcher, K. Dobson, M. S. Goodwin, H. Eydgahi, O. Wilder-Smith, D. Fernholz, Y. Kuboyama, E. B. Hedman, Ming-Zher Poh, and R. W. Picard, "iCalm: Wearable sensor and network architecture for wirelessly communicating and logging autonomic activity", *Information Technology in Biomedicine, IEEE Transactions on*, vol. 14, no. 2, pp. 215–223, 2010. 18

[70] Javad M. Akhlaghinia, "Occupancy monitoring and prediction in ambient intelligent environment", *PhD Thesis, School of Science and Technology, Nottingham Trent University*, 2010. 18, 21, 22, 23, 38

[71] R. Ali, M. ElHelw, L. Atallah, B. Lo, and Guang-Zhong Yang, "Pattern mining for routine behaviour discovery in pervasive healthcare environments", in *Information Technology and Applications in Biomedicine, 2008. ITAB 2008. International Conference on*, 2008, pp. 241–244. 18

[72] M. Ros, M. Delgado, and A. Vila, "A system to supervise behaviours using temporal and sensor information", in *Fuzzy Systems (FUZZ), 2010 IEEE International Conference on*, 2010, pp. 1–8. 18, 29

[73] Panagiota Kitsantas, Trent W. Moore, and David F. Sly, "Using classification trees to profile adolescent smoking behaviors", *Addictive Behaviors*, vol. 32, no. 1, pp. 9–23, 1 2007. 18

[74] Yu-Jin Hong, Ig-Jae Kim, Chul Ahn Sang, and Hyoung-Gon Kim, "Activity recognition using wearable sensors for elder care", in *2008 Second International Conference on Future Generation Communication and Networking (FGCN)*, Piscataway, NJ, USA, 13-15 Dec. 2008 2008, Imaging Media Res. Center, Korea Inst. of Sci. Technol., Seoul, South Korea, vol. 2, pp. 302–5, IEEE. 18

[75] H. Madokoro, K. Honma, and K. Sato, "Classification of behavior patterns with trajectory analysis used for event site", in *Neural Networks (IJCNN), The 2012 International Joint Conference on*, 2012, pp. 1–8. 18

[76] N. Suzuki, K. Hirasawa, K. Tanaka, Y. Kobayashi, Y. Sato, and Y. Fujino, "Learning motion patterns and anomaly detection by human trajectory analysis", in *Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on*, 2007, pp. 498–503. 18

[77] S. Nishio, H. Okamoto, and N. Babaguchi, "Hierarchical anomality detection based on situation", in *Pattern Recognition (ICPR), 2010 20th International Conference on*, 2010, pp. 1108–1111. 18

[78] T. Martin, B. Majeed, Beum-Seuk Lee, and N. Clarke, "Fuzzy ambient intelligence for next generation telecare", in *Fuzzy Systems, 2006 IEEE International Conference on*, 2006, pp. 894–901. 18, 29, 124

[79] Jyh-Yeong Chang, Jia-Jye Shyu, and Chien-Wen Cho, "Fuzzy rule inference based human activity recognition", in *Control Applications, (CCA) and Intelligent Control, (ISIC), 2009 IEEE*, 2009, pp. 211–215. 18, 29

[80] B. S. Lee, T. P. Martin, N. P. Clarke, B. Majeed, and D. Nauck, "Dynamic daily-living patterns and association analyses in tele-care systems", in *Data Mining, 2004. ICDM '04. Fourth IEEE International Conference on*, 2004, pp. 447–450. 18, 29

[81] M. Ahad, J. K. Tan, H. S. Kim, and S. Ishikawa, "Human activity recognition: Various paradigms", 2008, pp. 1896–1901. 18

[82] Sung-Ihk Yang and Sung-Bae Cho, "Recognizing human activities from accelerometer and physiological sensors", in *Multisensor Fusion and Integration for Intelligent Systems, 2008. MFI 2008. IEEE International Conference on*, 2008, pp. 100–105. 18

[83] Zhelong Wang, Ming Jiang, Yaohua Hu, and Hongyi Li, "An incremental learning method based on probabilistic neural networks and adjustable fuzzy clustering for human activity recognition by using wearable sensors", *Information Technology in Biomedicine, IEEE Transactions on*, vol. 16, no. 4, pp. 691–699, 2012. 18, 28

[84] Jinliang Song, Tiejian Luo, and Su Chen, "Behavior pattern mining: Apply process mining technology to common event logs of information systems", in *Networking, Sensing and Control, 2008. ICNSC 2008. IEEE International Conference on*, 2008, pp. 1800–1805. 18

[85] Marek Z. Reformat and Seyed Koosha Golmohammadi, "Rule- and OWA-based semantic similarity for user profiling", *International Journal of Fuzzy Systems,*, vol. 12, no. 2, pp. 87–102, 2010. 19

[86] G. Ali, N. A. Shaikh, and Z. A. Shaikh, "Agent-based user-profiling model for behavior monitoring", in *Future Networks, 2009 International Conference on*, 2009, pp. 3–7. 19

[87] J. A. Iglesias, P. Angelov, A. Ledezma, and A. Sanchis, "Creating evolving user behavior profiles automatically", *Knowledge and Data Engineering, IEEE Transactions on*, vol. 24, no. 5, pp. 854–867, 2012. 20

[88] Danni Wang, Clifford C. Federspiel, and Francis Rubinstein, "Modeling occupancy in single person offices", *Energy and Buildings*, vol. 37, no. 2, pp. 121–126, 2 2005. 20

[89] Dirk Saelens, Wout Parys, and Ruben Baetens, "Energy and comfort performance of thermally activated building systems including occupant behavior", *Building and Environment*, vol. 46, no. 4, pp. 835–848, 4 2011. 20

[90] Salman Taherian, Marcelo Pias, George Coulouris, and Jon Crowcroft, "Profiling energy use in households and office spaces", in *Proceedings of the 1st International Conference on Energy-Efficient Computing and Networking*, New York, NY, USA, 2010, e-Energy '10, pp. 21–30, ACM. 20

[91] Yong Joon Lee, Jun Wook Lee, Duck Jin Chai, Bu Hyun Hwang, and Keun Ho Ryu, "Mining temporal interval relational rules from temporal data", *Journal of Systems and Software*, vol. 82, no. 1, pp. 155–167, 1 2009. 20, 21

[92] James F. Allen, "Maintaining knowledge about temporal intervals", *Commun. ACM*, vol. 26, no. 11, pp. 832–843, 11 1983. 20

[93] T.L.M. van Kasteren, "Activity recognition for health monitoring elderly using temporal probabilistic models", *PhD Thesis: Promotor: F.C.A. Groen, Co-promotor: B.J.A. Krse, University of Amsterdam*, 2011. 21, 22, 23, 25

[94] M. Javad Akhlaghinia, A. Lotfi, C. Langensiepen, and N. Sherkat, "Occupancy monitoring in intelligent environment through integrated wireless

localizing agents", in *2009 IEEE Symposium on Intelligent Agents*, Piscataway, NJ, USA, 30 March-2 April 2009 2009, Sch. of Comput. Inf., Nottingham Trent Univ., Nottingham, UK, p. 7, IEEE. 21, 23

[95] A. Ahmad and F. Bait-Shiginah, "A nonconventional approach to generating efficient binary gray code sequences", *Potentials, IEEE*, vol. 31, no. 3, pp. 16–19, 2012. 23

[96] E. Agrell, J. Lassing, E. G. Strom, and T. Ottosson, "On the optimality of the binary reflected gray code", *Information Theory, IEEE Transactions on*, vol. 50, no. 12, pp. 3170–3182, 2004. 23

[97] Philomena M. Bluyssen, "Towards new methods and ways to create healthy and comfortable buildings", *Building and Environment*, vol. 45, no. 4, pp. 808–818, 4 2010. 23

[98] Pablo Bermejo, Luis Redondo, Luis de la Ossa, Daniel Rodrguez, Julia Flores, Carmen Urea, Jos A. Gmez, and Jos M. Puerta, "Design and simulation of a thermal comfort adaptive system based on fuzzy logic and on-line learning", *Energy and Buildings*, vol. 49, pp. 367–379, 6 2012. 23, 29

[99] P. M. Ferreira, S. M. Silva, A. E. Ruano, A. T. Negrier, and E. Z. E. Conceicao, "Neural network PMV estimation for model-based predictive control of HVAC systems", in *Neural Networks (IJCNN), The 2012 International Joint Conference on*, 2012, pp. 1–8. 23, 30

[100] Raad Z. Homod, Khairul Salleh Mohamed Sahari, Haider A. F. Almurib, and Farrukh Hafiz Nagi, "RLF and TS fuzzy model identification of indoor thermal comfort based on pmv/ppd", *Building and Environment*, vol. 49, pp. 141–153, 3 2012. 23, 29

[101] A. I. Dounis, P. Tiropanis, A. Argiriou, and A. Diamantis, "Intelligent control system for reconciliation of the energy savings with comfort in buildings using soft computing techniques", *Energy and Buildings*, vol. 43, no. 1, pp. 66–74, 1 2011. 23, 29

[102] Jin Woo Moon, Sung Kwon Jung, Youngchul Kim, and Seung-Hoon Han, "Comparative study of artificial intelligence-based building thermal control methods an application of fuzzy, adaptive neuro-fuzzy inference system, and artificial neural network", *Applied Thermal Engineering*, vol. 31, no. 14-15, pp. 2422–2429, 10 2011. 23, 30, 31

[103] Hani Hagras, Victor Callaghan, Martin Colley, and Graham Clarke, "A hierarchical fuzzygenetic multi-agent architecture for intelligent buildings online learning, adaptation and control", *Information Sciences*, vol. 150, no. 12, pp. 33–57, 3 2003. 23

[104] Wen-Tsai Sung and Yao-Chi Hsu, "Designing an industrial real-time measurement and monitoring system based on embedded system and zigbee", *Expert Systems with Applications*, vol. 38, no. 4, pp. 4522–4529, 4 2011. 23, 26

[105] Won-Suk Jang, William M. Healy, and Mirosaw J. Skibniewski, "Wireless sensor networks as part of a web-based building environmental monitoring system", *Automation in Construction*, vol. 17, no. 6, pp. 729–736, 8 2008. 23, 24

[106] Paul Davidsson and Magnus Boman, "Distributed monitoring and control of office buildings by embedded agents", *Information Sciences*, vol. 171, no. 4, pp. 293–307, 5/13 2005. 23

[107] "Smart energy starts here - echelon", 2013. 24

[108] Ali M. Malkawi and Ravi S. Srinivasan, "A new paradigm for human-building interaction: the use of CFD and augmented reality", *Automation in Construction*, vol. 14, no. 1, pp. 71 – 84, 2005. 24

[109] Cheng-Hung Tsai, Ying-Wen Bai, Chun-An Chu, Chih-Yu Chung, and Ming-Bo Lin, "PIR-sensor-based lighting device with ultra-low standby power consumption", *Consumer Electronics, IEEE Transactions on*, vol. 57, no. 3, pp. 1157–1164, 2011. 24

[110] Y. Chuo, M. Marzencki, B. Hung, C. Jaggernauth, K. Tavakolian, P. Lin, and B. Kaminska, "Mechanically flexible wireless multisensor platform for

human physical activity and vitals monitoring", *Biomedical Circuits and Systems, IEEE Transactions on*, vol. 4, no. 5, pp. 281–294, 2010. 25

[111] D. I. Tapia, S. Rodriguez, J. Bajo, J. M. Corchado, and O. Garcia, "Wireless sensor networks for data acquisition and information fusion: A case study", in *Information Fusion (FUSION), 2010 13th Conference on*, 2010, pp. 1–8. 25

[112] Wan Zhong-Bao, Xiao Yao, Wang Qing, and Gao Zhi-Gang, "WSN nodes for traffic data-acquisition system", in *Advanced Computer Control (ICACC), 2010 2nd International Conference on*, 2010, vol. 2, pp. 142–146. 25

[113] C. Gomez and J. Paradells, "Wireless home automation networks: A survey of architectures and technologies", *Communications Magazine, IEEE*, vol. 48, no. 6, pp. 92–101, 2010. 25, 26, 27, 40

[114] Chia-Ming Wu, Ruay-Shiung Chang, and Pu-I Lee, "An innovative scheme for increasing connectivity in zigbee networks", in *Parallel Processing (ICPP), 2011 International Conference on*, 2011, pp. 99–104. 26

[115] Fangsuo Wu, Xiaogang Hu, and Jianhua Shen, "Design and implementation of a low power wireless sensor network data acquisition system", in *Intelligent Computation Technology and Automation, 2009. ICICTA '09. Second International Conference on*, 2009, vol. 3, pp. 117–122. 26

[116] Paolo Baronti, Prashant Pillai, Vince W. C. Chook, Stefano Chessa, Alberto Gotta, and Y. Fun Hu, "Wireless sensor networks: A survey on the state of the art and the 802.15.4 and zigbee standards", *Computer Communications*, vol. 30, no. 7, pp. 1655–1695, 5/26 2007. 26

[117] Mitsugu Terada, "Application of zigbee sensor network to data acquisition and monitoring", *Measurement Science Review*, vol. 9, no. 6, 2009. 26

[118] Xinrong Zhang and Bo Chang, "Research of temperature and humidity monitoring system based on WSN and fuzzy control", in *Electronics and Optoelectronics (ICEOE), 2011 International Conference on*, 2011, vol. 4, pp. 300–303. 26

[119] J. Ohtsuka, T. Isokawa, A. Saitoh, N. Kamiura, N. Matsui, H. Kawa, and S. Nakashima, "Implementing localization scheme on zigbee wireless sensor network", in *SICE Annual Conference 2010, Proceedings of*, 2010, pp. 1195–1199. 26, 40

[120] Liu Yanfei, Wang Cheng, Qiao Xiaojun, Zhang Yunhe, and Yu chengbo, "An improved design of zigbee wireless sensor network", in *Computer Science and Information Technology, 2009. ICCSIT 2009. 2nd IEEE International Conference on*, 2009, pp. 515–518. 26

[121] Dae-Man Han and Jae-Hyun Lim, "Design and implementation of smart home energy management systems based on zigbee", *Consumer Electronics, IEEE Transactions on*, vol. 56, no. 3, pp. 1417–1425, 2010. 27

[122] V. Jakkula and D. Cook, "Outlier detection in smart environment structured power datasets", in *Intelligent Environments (IE), 2010 Sixth International Conference on*, 2010, pp. 29–33. 28

[123] Jinsung Byun and Sehyun Park, "Development of a self-adapting intelligent system for building energy saving and context-aware smart services", *Consumer Electronics, IEEE Transactions on*, vol. 57, no. 1, pp. 90–98, 2011. 28

[124] Yaohui Wang, Hongjian Chu, and Zhaoyang Qu, "The design of database anomalous detection model based on user behavior profile mining", in *Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on*, 2010, vol. 6, pp. 472–475. 28

[125] M. J. Akhlaghinia, A. Lotfi, and C. Langensiepen, "Soft computing prediction techniques in ambient intelligence environments", in *Fuzzy Systems Conference, 2007. FUZZ-IEEE 2007. IEEE International*, 2007, pp. 1–6. 29

[126] M. Tosun, K. Dincer, and S. Baskaya, "Rule-based mamdani-type fuzzy modelling of thermal performance of multi-layer precast concrete panels used in residential buildings in Turkey", *Expert Systems with Applications*, vol. 38, no. 5, pp. 5553–5560, 5 2011. 29

[127] G. Jahedi and M. M. Ardehali, “Genetic algorithm-based fuzzy-PID control methodologies for enhancement of energy efficiency of a dynamic energy system”, *Energy Conversion and Management*, vol. 52, no. 1, pp. 725–732, 1 2011. 29

[128] Benjamin Paris, Julien Eynard, Stphane Grieu, and Monique Polit, “Hybrid PID-fuzzy control scheme for managing energy resources in buildings”, *Applied Soft Computing*, vol. 11, no. 8, pp. 5068–5080, 12 2011. 29

[129] R. Karunakaran, S. Iniyan, and Ranko Goic, “Energy efficient fuzzy based combined variable refrigerant volume and variable air volume air conditioning system for buildings”, *Applied Energy*, vol. 87, no. 4, pp. 1158–1175, 4 2010. 29

[130] Chengli Li, Qihui Zhang, and Tingting Mou, “The study of neural network in the application of PMV index”, in *System Science, Engineering Design and Manufacturing Informatization (ICSEM), 2010 International Conference on*, 2010, vol. 1, pp. 289–292. 30

[131] Jian Liang and Ruxu Du, “Thermal comfort control based on neural network for HVAC application”, in *Control Applications, 2005. CCA 2005. Proceedings of 2005 IEEE Conference on*, 2005, pp. 819–824. 30

[132] M. Castilla, J. D. lvarez, M. G. Ortega, and M. R. Arahal, “Neural network and polynomial approximated thermal comfort models for hvac systems”, *Building and Environment*. 30

[133] S. Jassar, Z. Liao, and L. Zhao, “Adaptive neuro-fuzzy based inferential sensor model for estimating the average air temperature in space heating systems”, *Building and Environment*, vol. 44, no. 8, pp. 1609–1616, 2009. 30, 31

[134] S. Ari, H. E. Khalifa, J. F. Dannenhoffer, P. Wilcoxen, and C. Isik, “Fuzzy logic and neural network approximation to indoor comfort and energy optimization”, in *Fuzzy Information Processing Society, 2006. NAFIPS 2006. Annual meeting of the North American*, 2006, pp. 692–695. 31

[135] Pico Technology, “Picolog 1000 series user’s guide”. 39

[136] Dae-Man Han and Jae-Hyun Lim, "Design and implementation of smart home energy management systems based on zigbee", *Consumer Electronics, IEEE Transactions on*, vol. 56, no. 3, pp. 1417–1425, 2010. 40

[137] Martin Hebel, George Bricker, and Daniel Harris, *Getting Started with XBee RF Modules - A Tutorial for BASIC Stamp and Propeller Microcontrollers*, PARALLAX Inc. 40

[138] Fred Eady, *Hands On ZigBee - Implementing 802.15.4 with Microcontrollers*, Elsevier Inc., Linacre House, Jordan Hill, Oxford OX2 8DP,UK, 2007. 41

[139] Matlab and Simulink, "Stateflow 7 user's guide", *The MathWorks, Inc.*, 2010. 53

[140] K.A. Stroud, *Engineering Mathematics*, Maccmillan Press Ltd., London, 1995. 58

[141] Seung-Seok Choi, Sung-Hyuk Cha, and Charles C. Tappert, "A survey of binary similarity and distance measures", *Journal of Systemics, Cybernetics and Informatics*, vol. 8, no. 1, pp. 43–48, 5 2010. 61, 69

[142] B. Zhang and S.N. Srihari, "Properties of binary vector dissimilarity measures", *Proc. JCIS Int'l Conf. Computer Vision, Pattern Recognition, and Image Processing*, 9 2003. 61

[143] Sung-Hyuk Cha, Charles Tappert, and Sungsoo Yoon, "Enhancing binary feature vector similarity measures", *Journal of Patteren Recognition Research*, vol. 1, pp. 63–77, 2006. 62, 68

[144] B. G. Mujtaba, "Ethical implication of employee monitoring: What leaders should consider", *Journal of Applied Management and Entrepreneurship*, vol. 18, no. 3, 7 2003. 62

[145] Johnathan Yerby, "Legal and ethical issues of employee monitoring", *Journal of Applied Knowledge Management*, vol. 1, no. 2, pp. 44–55, 2013. 62

[146] Kilem Li Gwet, *Handbook of inter-rater reliability. 3nd ed.*, Advanced Analytics, LLC; USA, Advanced Analytics, LLC, P.O. Box 2696,Gaithersburg, MD 20886-2696,USA, 2012. 64, 65

[147] S. M. Pincus, "Approximate entropy as measure of system complexity", *Poceedings of the National Academy of Sciences of the United States of America*, vol. 88, no. 6, pp. 2297–2301, 1991. 65, 67

[148] Ki Chon, C. Scully, and Sheng Lu, "Approximate entropy for all signals", *Engineering in Medicine and Biology Magazine, IEEE*, vol. 28, no. 6, pp. 18–23, 2009. 65

[149] M. Sattler, R. Wilson, T. El-mezyani, S. Srivastava, C. S. Edrington, and D. Cartes, "Approximate entropy measure of nonlinearities in a AC/DC converter", in *Proceedings of the 2010 Conference on Grand Challenges in Modeling & Simulation*, Vista, CA, 2010, GCMS '10, pp. 219–224, Society for Modeling & Simulation International. 66

[150] Sheng Lu, Xinnian Chen, J. K. Kanters, I. C. Solomon, and K. H. Chon, "Automatic selection of the threshold value for approximate entropy", *Biomedical Engineering, IEEE Transactions on*, vol. 55, no. 8, pp. 1966–1972, 2008. 66, 90

[151] Minghui Jiang, "A linear-time algorithm for hamming distance with shifts", *Theory of computing System*, vol. 44, no. 3, pp. 349–355, 2009. 68

[152] Vladimir A. Kulyukin Abraham Bookstein and Timo Raita, "Generalized hamming distance", *Journal of Information Retrieval*, vol. 5, no. 4, pp. 353–375, 2002. 68, 69, 70

[153] Minghui Jiang, "Linear-time algorithm for hamming distance with shifts", *Theory of Computing Systemsl*, vol. 44, no. 3, pp. 349–355, 2009. 69

[154] Paolo Tormene, Toni Giorgino, Silvana Quaglini, and Mario Stepanelli, "Matching incomplete time series with dynamic time warping: an algorithm and an application to post-stroke rehabilitation", *Artificial Intelligence in Medicine*, vol. 45, pp. 11–34, 2009. 70, 72

[155] Stan Salvador and Philip Chan, "Fastdtw: Toward accurate dynamic time warping in linear time and space", *KDD Workshop on Mining Temporal and Sequential Data*, pp. 70–80, 2004. 70

[156] Ming Hsiao Ko, Geoff West, Svetha Venkatesh, and Mohan Kumar, "Using dynamic time warping for online temporal fusion in multisensor systems", *Information Fusion*, vol. 9, no. 3, pp. 370–388, 7 2008. 70, 71, 108

[157] Pavel Senin, *Dynamic Time Warping Algorithm Review*, Information and Computer Science Department,University of Hawaii at Manoa,Honolulu, USA, 2008. 72

[158] Mirna Safi, "The immigrant intergration process in france: Inequalities and segmentation", 2008, vol. 49, pp. 3–44. 89, 144

[159] Sheng Lu, Xinnian Chen, J. K. Kanters, I. C. Solomon, and K. H. Chon, "Automatic selection of the threshold value for approximate entropy", *Biomedical Engineering, IEEE Transactions on*, vol. 55, no. 8, pp. 1966–1972, 2008. 91

[160] Kalon K. L. Ho, George B. Moody, Chung-Kang Peng, Joseph E. Mietus, Martin G. Larson, Daniel Levy, and Ary L. Goldberger, "Predicting survival in heart failure case and control subjects by use of fully automated methods for deriving nonlinear and conventional indices of heart rate dynamics", 1997. 91

[161] Young Sook Son and Jangsun Baek, "A modified correlation coefficient based similarity measure for clustering time-course gene expression data", *Pattern Recognition Letters*, vol. 29, no. 3, pp. 232–242, 2/1 2008. 99

[162] Piotr Zerzucha and Beata Walczak, "Concept of (dis)similarity in data analysis", *TrAC Trends in Analytical Chemistry*, vol. 38, pp. 116–128, 9 2012. 101

[163] Taiwo Adetola Ojurongbe, "Comparison of different proximity measures and classification method for binary data", *PhD Thesis, Faculty of Agriculture Sciences, Nutritional Sciences and Environmental Management, Justus Liebig University Gieen*, 2012. 101

[164] M. Daszykowski, R. Danielsson, and B. Walczak, "No-alignment-strategies for exploring a set of two-way data tables obtained from capillary electrophoresis-mass spectrometry", *Journal of Chromatography A*, vol. 1192, no. 1, pp. 157–165, 5/23 2008. 101, 103

[165] Weikai Yan and Nicholas A. Tinker, "Biplot analysis of multi-environment trial data: Principles and applications", *Canadian Journal of Plant Science*, vol. 86, pp. 623–645, 2/26 2006. 101, 104

[166] Jesus Martin-Rodriquez, Purificacion Galindo Villardon, and Jose L. Vicente Villardon, "Comparison and integration of subspaces from a biplot perspective", *Journal of Statistical Planning and Inference*, vol. 102, no. 2, pp. 411–423, 4/1 2002. 101

[167] Eamonn Keogh, "Exact indexing of dynamic time warping", in *Proceedings of the 28th international conference on Very Large Data Bases*. 2002, VLDB '02, pp. 406–417, VLDB Endowment. 108, 113

[168] Akin Avci, Stephan Bosch, Mihai Marin-Perianu, Raluca Marin-Perianu, and Paul Havinga, "Activity recognition using inertial sensing for healthcare, wellbeing and sports applications: A survey", in *Architecture of Computing Systems (ARCS), 2010 23rd International Conference on*, 2010, pp. 1–10. 126, 168

[169] G. S. Chambers, S. Venkatesh, G. A. W. West, and H. H. Bui, "Hierarchical recognition of intentional human gestures for sports video annotation", in *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, 2002, vol. 2, pp. 1082–1085. 126, 168

[170] H. Seki, "Fuzzy inference based non-daily behavior pattern detection for elderly people monitoring system", in *Engineering in Medicine and Biology Society 2009 Annual International Conference of the IEEE*, 2009, pp. 6187–6192. 131

[171] "MIT media lab: Affective computing group", 2013. 147