Trend Analysis for Human Activities Recognition

ABUBAKER MOHAMED ELBAYOUDI

School of Science and Technology

A thesis submitted in partial fulfilment of the requirements of Nottingham Trent University for the degree of

Doctor of Philosophy

January 2018

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 is carried out at Nottingham Trent University between February 2014 and January 2018, while the researcher was working as a full-time doctoral research student.

I owe my utmost sincere gratitude to my honourable director of studies, Prof. Ahmad Lotfi, for his guidance over the last four years. His heart and his office always open for me at any time despite his many responsibilities. Also, his relentless effort, extensive suggestions and valuable comments to guide me completing my work were indispensable to the accomplishment of this research. I would like to extend my sincere thanks to Dr. Caroline Langensiepen for her valuable advice, support, feedback and knowledge throughout the development of my thesis. Also, I would like to thank Dr. Kofi Appiah for his support and feedback throughout the course of this project work. Furthermore, I really appreciate and thank my sponsor, Ministry of Higher Education of Libya for awarding me this Scholarship and the cultural affairs attach at the Libyan embassy in London.

I would like to pay tribute to my family for their support, sympathy, and patience throughout the journey of this study, particularly, I would like to say thanks to my wife Zaineb and our lovely boys, Mohamed, Yamen and Aban. Your joy and love have always put me up during the well-being and sadness of the last four years.

My thanks are extended to my father's soul who always encouraged me to carry on in my study, also I thank my mother, sisters, brothers, brothers in law and uncles. Their prayers and faith have inspired and encouraged me to continue with this study. I owe a huge debt of gratitude to my brother Khalid and my brother in law Younes for being my guarantor underwriting contract with my sponsor and they were looking after my business back in Libya. Thanks also should be gone to friends and colleagues in the School of Science and Technology at Nottingham Trent University.

Abubaker Mohamed Elbayoudi
January 2018

Abstract

Smart environments equipped with appropriate sensory devices are used to measure people's activities. These activities represent Activities of Daily Living (ADL) or Activities of Daily Working (ADW). Measuring progressive changes in activities is a subject of research interest. A number of medical conditions and their treatments are associated with progressive changes such as reduced movement over time.

The aim of this research is to determine means of inspecting trends in the ADL/ADW to identify progressive changes and predict behavioural abnormalities. The ADL/ADW pattern will change over time and this is a consequence of the individual's condition. Identifying evolving behavioural patterns will help to predict the trend in the ADL/ADW behavioural pattern before any abnormalities are identified. The data provided for this investigation are from real environments (home and office). Additionally, a simulator is developed to generate simulated data for ADLs.

To answer the research question identified in this research, the initial investigation was conducted and a novel **Human Behaviour Momentum Indicator** (HBMI) is proposed. The HBMI is introduced to identify changes based on activities recorded from a single sensor. To show the effectiveness of the proposed approach, results are compared with Relative Strength Index (RSI). The results show that trends in ADL or ADW can be detected and the direction of the activity's trend is predicted.

To represent a holistic report based on a multiple sensors/activities representing progressive changes in the participant's behaviour, a novel

Human Behaviour Indicator (HBI) is also proposed. The proposed HBI indicator is constructed as a composite indicator, which will compute progressive changes in behaviour based on the events that are performed during the entire day. The percentage of changes between events is used to compare events and measure the progressive changes. The proposed technique identifies the user's daily behaviour and distinguishes between normal and abnormal behavioural patterns of the ADLs or ADWs. Analysis of the data indicates that the HBI could clearly differentiate between the normal and the abnormal behaviour and give a warning status with a confidence level.

Identifying trends in ADLs or ADWs using trend analysis techniques are investigated to interpret the behavioural changes in a suitable format to be understood by the carers or supervisors.

Publications

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

Abubaker Elbayoudi, Ahmad Lotfi, Caroline Langensiepen, The Human Behaviour Indicator; A Measure of Behavioural Evolution, Expert Systems with Applications (submitted).

Refereed Conference Papers

Abubaker Elbayoudi, Ahmad Lotfi, Caroline Langensiepen, Kofi Appiah, Trend Analysis Techniques in Forecasting Human Behaviour Evolution, The tenth International Conference on PErvasive Technologies Related to Assistive Environments, Rhodes, Greece, pp. 293-299,
DOI:10.1145/3056540.3076198, 2017.

Abubaker Elbayoudi, Ahmad Lotfi, Caroline Langensiepen, Kofi Appiah, Determining behavioural trends in an ambient intelligence environment, The ninth International Conference on PErvasive Technologies Related to Assistive Environments, Corfu, Greece, pp. 31-38,
DOI:10.1145/2910674.2935834, 2016.

Abubaker Elbayoudi, Ahmad Lotfi, Caroline Langensiepen, Kofi Appiah, Modelling and simulation of activities of daily living representing an older adult's behaviour, The eightieth International Conference on PErvasive Technologies Related to Assistive Environments, Corfu, Greece, pp. 67-74, DOI:10.1145/2769493.2769544, 2015.

Ahmad Lotfi, Caroline Langensiepen, Abubaker Elbayoudi, Interpretation of behaviour evolution in activities of daily living, The eightieth International Conference on PErvasive Technologies Related to Assistive Environments, Corfu, Greece, pp. 84-85,
DOI:10.1145/2769493.2769514, 2015.

Contents

Nomenclature

Acronyms

ADL	Activities of Daily Living
ADW	Activities of Daily Working
AmI	Ambient Intelligence
ANFIS	Adaptive-Network based Fuzzy Inference System
ANN	Artificial Neural Networks
APMA	Activity Prediction Moving Average
CI	Computational Intelligence
CMA	Cumulative Moving Average
DSMC	Direct Simulation Monte Carlo
EW	Equal Weighting
EWHBI	Exponential Weighted Human Behaviour Indicator
EWMA	Exponentially Weighted Moving Average
GMM	Gaussian Mixture Model
HAR	Human Activity Recognition
HBI	Human Behaviour Indicator
HBMI	Human Behaviour Momentum Indicator
HMM	Hidden Markov Model
HHMM	Hierarchical Hidden Markov Model
ICT	Information and Communication Technologies
MA	Moving Average
MAR	Missing at Random
MACD	Moving Average Convergence/Divergence
MCI	Mild Cognitive Impairment

MKT	Mann-Kendall Test
MMA	Modified Moving Average
NMAR	Not Missing at Random
PD	Parkinson's Disease
RMSE	Root Mean Square Error
RSI	Relative Strength Index
SAPM	Sensors Activity Pattern Matching
SCI	Synthetic Composite Indicator
SKT	Seasonal Kendall Test
SMA	Simple Moving Average
SMM	Simple Moving Median
SVM	Support Vector Machine
WHBI	Weighted Human Behaviour Indicator
xED	Extended Episode Discovery **Greek Symbols**
τ	Measures the correlation between a chosen parameter and time
σ	Calculate variance
θ	The number of tied groups
α	The critical value that is found using a Chi-square distribution table

List of Figures

List of Tables

Chapter 1

Introduction

It is always useful to find means of improving lifestyles by monitoring the progressive changes in people's activities of daily living (ADL) or activities of daily working (ADW). The number of older adults who need peripheral help in their daily activities is increasing and it is predicted to be approximately 22% of the world population by 2050 [1]. This will lead to some problems in caring systems for those people and increase the number of care providers is not a realistic solution. It is suggested that smart environments technologies can help in solving some of the related problems [2]. An office worker could face major causes of stress that could affect their well-being or behaviour because of the lack of control of their environmental conditions. Therefore, using current technology in the smart office to manage the performance of the office's environment such as lighting, heating and Personal Computer so as to be more responsive to the user's habits and routines, would be more acceptable and better for their users' wellbeing. [3, 4, 5].

Monitoring people in their smart environments while they are carrying out their activities and inferring decisions as of when to intervene is an open area for research [6]. The ADL or ADW may all be affected by changes in a person's health or well-being. For example, total sleeping duration could follow a consistent pattern unless there are step changes due to some medical conditions [7, 8].

It is necessary to integrate pervasive technologies into work and home environments to collect the data that represents ADL and ADW. Identifying trends in users' behaviour over a long period will help to forecast abnormalities in ADL and ADW. In this thesis, progressive changes in people's behaviour are referred to as **behavioural evolution**. For instance, identifying the progressive changes in the behaviour of a person suffering from mild cognitive impairment will allow the caregiver to monitor and intervene when abnormal behaviour is predicted. The ADL pattern may change over time, and it is important to forecast the trend in the ADL pattern before any abnormalities are identified.

Smart environments including smart homes and smart offices should be able to detect, track and recognise participants in their places. Smart environments may include the following aspects [9]:

- The number of participants in the smart environment,
- The participant's recognition,
- The participant's activities, and
- The participant's movements between different areas.

It is a very important factor in designing a smart environment to ensure that the technology does not interfere, with participants' activities and all sensory devices in the data collection network should operate autonomously [10]. The sensory devices at home could include house electrical devices e.g. cooker and fridge, domestic objects e.g. taps, bed and sofa, temperature conditioning devices e.g. air conditioning and radiator, personal computers and chairs. These devices can be activated automatically or monitored remotely. Some potential advantages of this technology are:

- By monitoring the progressive changes in the participant's behaviour or current activities, safety concerns can be raised and ensure help is provided whenever a possible abnormal behaviour is observed,
- It can adapt and control the devices that make the participant more comfortable, such as changing the temperature automatically or controlling the use of the electricity to reduce cost and for energy efficiency.

The aim of this research is to identify means of inspecting trends in the ADL or ADW to identify progressive changes gathered from a sensor network forming an Ambient Intelligence (AmI) environment [1]. The smart environments should be equipped with a sensor network to monitor the participants' behaviour and identify their ADL or ADW. The data used in this research are collected using low cost and readily available sensors, which can be installed by the participants themselves.

The rest of this chapter is organised in the following order. An overview of this research is presented initially followed by the aims of this thesis and the proposed objectives presented in Section 1.2. Section 1.3 introduces the major contribution of the thesis. Finally, Section 1.4 has the outline of the remaining chapters of this thesis.

1.1 Overview of the Research

In an AmI environment, the participant's behaviour can be extracted from a sensory network which represents the individual participant's activities model. Statistical methods such as Bayesian networks [11, 12] and Hidden Markov Models (HMMs) [13, 14] are used in many researches to model the behaviour of people who live or work in an intelligent environment. There are many challenges in using such statistical methods to find the relationship between the sensory data and identifying the actual behaviour of the participant in an intelligent environment. It is very difficult to model large low-level sensory data sets [6], because of the complexity of the outputs from these methods [15].

Understanding human behaviour from low level sensory data, interpreting and summarising a large amount of data and presenting it in an understandable format are the main challenges that the research here faced. These challenges could be tackled using different techniques such as Computational Intelligence techniques or statistical methods integrated with sensory data.

The main research question addressed in this thesis is to investigate the use

[1]The terms "Ambient Intelligence (AmI)", "Intelligent Environment" and "Smart Environment" are used interchangeably. Examples of such environments are intelligent or smart homes and intelligent or smart offices.

of sensors network technology to monitor humans in smart environments and analyse extracted data that represents the participant's behaviour. In particular, this study is trying to answer the following questions:

- How to extract behavioural patterns of a person's ADL or ADW by analysing sensory data collected from different sensors in a smart environment.
- How the sensory data is processed to become time series then identify trends within the data.
- How the data is forecast to extract important daily patterns from them and predict the direction of the trends in the data.
- How the multiple data that represent multiple activities or events in every single day are combined in one single datum to represent the participant's overall behaviour.
- How unexpected patterns and anomalies within the combined data can be identified.
- How to validate and test the proposed solutions on data collected from real environments.

To answer the above questions, the following aims and objectives are identified for this research.

1.2 Aims and Objectives

The availability of sensor network technology will make it possible to support and help people in their life by monitoring their activities in smart environments. However, using various sensors to collect a large amount of data representing human activities and translate it into a meaningful knowledge is a challenging task. Summarising extracted data of a person using a smart environment to understand the progressive changes of this person's behaviour or predict his/her future behaviour status is a research challenge due to the complexity of human behaviour [16, 17].

The research aims to identify trends in ADL or ADW and interpret them in a suitable output format. Change of behaviour of a participant in an AmI environment is an indicator of this person's social and health status. The project is primarily concerned with the interpretation of progressive changes in the participant's behaviour.

This research tries to find an acceptable solution to combine the ADLs or ADWs of a person who lives or works independently in his/her own home or office on daily basis and present it in one single value for each day. Most people would prefer to use non-intrusive sensor technology [18, 19]. Non-intrusive sensors will not affect their normal activities whereas other sensors may need cooperation from the user. For instance, wearable sensors need to be worn and sometimes the participant might have forgotten to wear them [20]. This research will provide a holistic view of the monitored person's behaviour and it will forecast the behaviour evolution of the participant and it will raise an early warning of an abnormal behaviour when it is observed or expected to happen in near future.

The present research, therefore, will employ and assess such approaches to discover suitable methods for constructing Synthetic Composite Indicators (SCI). In particular, it aims to deliver an intelligent technique as a framework to build a Human Behaviour Indicator (HBI) that can help enable consistent and transparent assessments and forecasting of the progressive changes in human behaviour. The HBI represents a holistic report based on multiple sensors/activities representing progressive changes in the participant's behaviour. As a starting point for developing HBI, the investigation initially focused on a single sensor. A novel Human Behaviour Momentum Indicator (HBMI) is introduced to identify changes based on activities recorded from a single sensor. The proposed methods, HBMI and HBI, will be applied to several data sets representing ADLs or ADWs. They are evaluated using real data sets and the results show that trends in ADL or ADW can be detected and the direction of the activity's trend is predicted.

To accomplish the aim of this research, the following objectives are identified:

- To investigate different methods for representing big sensory data sets which representing the participants' behavioural patterns.

- To investigate existing trend analysis techniques and their suitability for measuring trends in ADL or ADW.

- To compare the performance of different techniques to assess the most appropriate technique of understanding and interpretation of the sensory data sets collected from smart environments.
- To investigate alternative methods for measure human behaviour evolution and to introduce a novel HBMI to be used as an intelligent technique that can measure progressive changes in human behaviour and can predict the direction of trends in the behaviour.
- To construct a novel HBI technique to measure the overall changes in human behaviour and give a holistic overview of the human behaviour evolution.
- To validate the effectiveness of the introduced HBMI and HBI frameworks and to evaluate their strengths and weaknesses. Several real data sets from different smart environments are used as case studies.

1.3 Research Contribution

The main contributions of this thesis are:

- To identify the behavioural pattern of a participant who lives or works in a smart environment using sensory data.
- To investigate techniques that can visualise large binary data sets in a suitable form. For example, start time, duration and disturbances of occupying one place on daily basis can be used for modelling the occupancy pattern of this place and it may then be used to forecast the behaviour.
- To investigate different techniques that can be applied to determine trends in human behaviour. For instance, using one of the versions of moving average to determine the trend in a data set that represents sleeping behaviour.
- To determine outliers or anomalies within the behavioural patterns of the behaviour of a participant who uses a smart environment. The identification is based on different trend analysis techniques.

- To investigate techniques that can be used to aggregate the data of ADLs or ADWs and identify the trends and predict the trend direction of aggregated data. To provide an understandable holistic report of a monitored person, a novel HBMI is introduced. For example, using this approach to aggregate more than one feature (start time, duration, number of disturbances and end-time of sleeping) to represent the whole sleeping trend and predict the direction of the trend.

- To introduce a novel HBI technique that can be used to aggregate the events carried out by a participant in a smart environment. The novel approach that is investigated in this research is based on using statistical techniques to build a composite indicator that can aggregate the data and presents it in one understandable graph representing long term human behaviour evolution.

1.4 Thesis Outline

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

Chapter 2: Literature Review - This chapter discussed the relevant literature in the field of intelligent environments. The literature is covered the following areas: healthcare challenges, human behaviour recognition, trend analysis techniques, momentum indicators and composite indicators techniques, representation and abnormal detection using statistical methods and computational intelligence techniques. The literature gives an overview of the available technologies that are used for modelling the human behaviour.

Chapter 3: Data Preparation and Data Representation - This chapter describes the data sets representing the ADLs and ADWs for the people who occupied smart environments. The data preparation is also presented in this chapter. In addition, the sensory data are represented using different techniques. Some methods of interpretation and visualisation of data are discussed. These methods are implemented on our data to understand the behavioural pattern evolution of the participants.

Chapter 4: Trend Analysis and Prediction Techniques - This chapter provides an overview of trend analysis techniques that can be applied to human behaviour data sets. Then some existing prediction techniques which are used in time series prediction are reviewed. The chapter begins by presenting the trend analysis techniques such as different moving average techniques and seasonal Kendall test. In addition, novel prediction approaches based on trend analysis techniques are introduced as time series predictors. Then, the chapter introduces different momentum indicators methods used in this thesis and discusses their benefits in time series prediction.

Chapter 5: Applying Trend Analysis and Prediction Techniques for single Activity - In this chapter, some techniques that can be used in analysing binary sensory data to identify normal behaviour, distinguish any abnormalities and determine possible trends in the behaviour are presented. A novel momentum indicator that can be used to predict the future behaviour direction is proposed. This indicator can detect abnormal activities. The chapter starts with an overview of trend analysis techniques that are used to detect trends in human behaviour in this research. Then prediction techniques are described and novel ideas to predict the future of trends and behaviour are implemented. The chapter concludes that the proposed techniques to identify trends and predict the trend's direction can find trends direction and anomalies within the sensory data.

Chapter 6: The Human Behaviour Indicator - In this chapter, a novel composite indicator to aggregate sensory data is presented. The indicator is built to aggregate multivariate data features represents the events carried out by the participant in a smart environment. Different models of the composite indicator are tested and compared to find out the best model to aggregate the sensory data.

Chapter 7: Conclusions and Future Works - The conclusions of this thesis are presented in this chapter and some future research in progressive changes in human behaviour are proposed.

Chapter 2

Literature Review

2.1 Introduction

Progressive changes in the human behaviour can be monitored and measured using appropriate measurement devices and communication equipment that are available to support collecting data from smart environments. A network of sensors can be installed in a smart environment to continuously monitor the participants, which will help a supervisor to observe the progressive changes in the participant's behaviour. Many smart environment projects that allow a human to live or work independently have been developed. They can monitor the participants who may have or may have not disabilities [21, 22]. For instance, living in a nursing home is quite expensive; therefore, a smart home is a preferable living environment [23]. Smart home environments can perceive long-term changes that may cause health concerns. Such systems will alert carers and family of any significant changes in the occupant's behaviour, diet, daily tasks or health. Smart environments offer such solutions by using human behaviour recognition to monitor the person's activities and alert the carer if something abnormal is detected. In another example, to reduce energy consumption in a smart environment, human behaviour can be monitored to improve their habits in terms of reducing the energy consumption and carbon emissions [24]. Human behaviour recognition has been demonstrated to be a valuable key to understand people's needs [22].

The analysis of human behaviour using trend analysis techniques will show how people act in their daily life, therefore providing a better understanding of their necessities and demands. The analysis of human behaviour will help to understand people's habits and daily routines. So far, assisted living systems have attempted to meet the understanding of personal healthcare challenges when bringing healthcare through Information and Communication Technologies (ICT) [25].

This chapter reviews existing research studies on trends in human behaviour, human behaviour recognition, progressive changes in human behaviour, and abnormal detection algorithms and techniques. This chapter is structured as follows: in Section 2.2, the trend analysis for human behaviour is reviewed. Related literature on human behaviour recognition, Composite Indicators and abnormality detection are reviewed in Sections 2.3, 2.4 and 2.5 respectively. A methodology of the research is presented in Section 2.7. A summary is presented in Section 2.8.

2.2 Trend Analysis for Human Activities Recognition

Trend analysis for human behaviours (monitored in a smart environment) is investigated by some researchers in this field [6]. Mahmoud et. al. in [26] introduced the importance of trends in activities of daily living of a single elderly person. An algorithm called Extended Episode Discovery (xED) is introduced in [27] to search for regular patterns, highlighting the periodicity and variability of each discovered pattern. However, the data sets used in this research to validate their algorithms are very small.

Virone et. al. in [28] presented a pattern mining model to monitor high-level activities of an elderly person to help them live independently. The system will send an alert to caregivers about the evolution of the behaviour status of their monitored user. Their idea is good but they did not use their model on real data. Kemp et. al. in [29] introduced an algorithm to generate information and summarisation of changes in an elderly person's behaviour for long-term monitoring. They focus on reducing the number of transmissions required by a wearable monitoring system. The authors in [30] introduce a framework to determine the wellness of an elderly person. They use the double exponential smoothing strategy to determine the trends of the older adult's activity. However, they claim that the sequence pattern of the sensors on a day at a particular period could be predicted using Sensors Activity Pattern Matching (SAPM) technique, which is not possible for human behaviour.

A linguistic summarisation for describing long-term trends of change in human behaviour is presented in [31]. The procedure is to provide information to elder adults, carers and family in an understandable language by adapting a measure of similarity for comparing behaviours that are adapted over time. Their idea is based on clustering and comparing the similarity between the activities at a specific time to detect trends. However they concentrate on linguistic summarisation more than trend analysis itself. Researchers in [32] described models for detecting behavioural and health-related changes in a patient's behaviour who is monitored continuously in a smart home. They used Holt's linear trend method to predict different physiological parameters of the monitored user.

Many factors could cause cycles in data such seasonal climatic changes, tides, changes in vehicle traffic patterns during the day, and so on. These kinds of changes are not "trends" because they do not indicate long-term change. Therefore, the trend should be considered as a cycle with a rising/falling long-term level with random fluctuation about the cycle. The common trend analysis techniques are categorised into three main types as they are presented in Sections 2.2.1, 2.2.2 and 2.2.3 respectively.

2.2.1 Traditional Techniques

Mathematical techniques are used to forecast all types of trends. The exploration of data analysis is to identify the diverse types of trends in historical data [33]. The most common methods include various forms of weighted smoothing methods, turning point analysis, decomposition, adaptive filtering, Box-Jenkins, moving average and curve fitting analysis. These mathematical models have a common feature which is that historical data is the only criteria for producing a forecast. However, if two people are using the same model on the same data the forecasting may not be the same because mathematical models involve smoothing constants, coefficients and other parameters that must be adjusted by the user.

Many researchers use mathematical trend analysis techniques to identify disparities and to propose future goals. Authors in [34] have used moving average to find trends in the progress of reducing mortality among persons 65 to 74 years of age. Fitting curve analysis technique is used in [35] to study bones and similar materials such as ivory from prehistoric sites which have recorded a wealth of information on the past ways of life and its changes.

2.2.2 Statistical Techniques

Statistical trend analysis techniques are often used to extract an underlying pattern of behaviour in time series. Time series data could be partly or nearly completely hidden by noise and the statistical methods have the potential to get their trend. These techniques could be simply described as a trend estimation, which can be undertaken within a formal regression analysis.

Researchers have proposed many statistical techniques to detect trends [36, 37]. The methods are classified into parametric, non-parametric and mixed techniques. The non-parametric test do not require any further major assumptions about the data [38]. The parametric methods require data to be normally distributed. However, the assumption of independence is more important than the normality assumption. Most of the proposed methods give reasonable estimates of the trend even if the normality assumption is violated if they are provided with the true distribution which is not excessively distorted and the independence assumption holds. If the independence assumption is badly broken then the trend estimates will be quite poor [39].

The most common statistical methods to extract trends according to [39] are linear regression, Spearman's rank correlation coefficient, Non-seasonal Mann-Kendall test, Regional non-seasonal Mann-Kendall test, Seasonal Kendall test and Regional seasonal Kendall test.

In general, trend analysis techniques are widely used in many scientific studies. For instance, authors in [40] used Laplace Test for trend analysis and prediction of incipient faults for power systems. In healthcare studies, some researchers use trend analysis to monitor trends in disease and death rates and social and behavioural risk factors that may participate in adverse events [41].

2.2.3 Computational Intelligence Techniques

Many Computational Intelligence (CI) techniques are used to identify trends in data mainly in economic studies [42]. The most common methods used in trend analysis include neural networks, fuzzy logic and genetic algorithms. In addition, there is another approach based on hybrids of two or more CI techniques together or between CI and statistical techniques [43]. For instance, a multiple fuzzy inference system is used to predict daily trading decisions [44]. Chourmouziadis et. al. in [45] used fuzzy system rules along with moving average to assist investors in terms of trend analysis in short term stock trading. Neural networks are used for constructing trend impact analysis (TIA) in a stock market [46]. Atsalakis et. al. in [47] present a Neuro-Fuzzy based methodology to forecast short term stock trends during turbulent stock market periods. However, CI techniques still need more investigations in terms of trend analysis.

2.3 Human Behaviour Recognition

Human Behaviour Recognition or Human Activity Recognition (HAR) is a subject of interest for many researchers. The majority of existing researches in monitoring are concentrating on the detection of normal and abnormal behaviour and how to distinguish between these two groups of activities. Therefore, recognising significant changes in human behaviour early will help carers/supervisors to act to avoid prospective problems early as well.

2.3.1 AmI Applications to Support Monitors

Many people find monitoring of their older adults are very challenging, especially if they have dementia. There are many approaches introduced in human activity recognition. For example, Fahad et. al. in [48] perform monitoring of changes in a daily routine of a person lives in a smart home using long-term analysis.

Researchers in [49] discussed the importance of social and personality factors for building intelligent systems that can interact with individuals in such a way that would result in behavioural change. In other research SanMiguel et. al. proposed an automatic system to recognise human actions from video sequences taken in a room [50]. However, more investigations in applications to support monitors will bring better solutions to help them in their work.

Human activity recognition aims to get information about people's activities automatically using sensing modalities and analysing them [51]. In general, human activity understanding covers activity recognition and pattern discovery. The first deals with the detection of human activities based on a predefined activity model and activity pattern discovery are finding unknown patterns directly from low-level sensor data without any predefined models or assumptions. Both techniques are aimed to improve human activity technology and they are complementary to each other [52, 53]. Using activity recognition and pattern discovery in healthcare and assistance and wellness are possibly the fields that most actively leverage the knowledge gained from the analysis of human behaviour. Examples of how healthcare benefits from this technology include supporting healthier lifestyles e.g., encouraging exercising [54], preventing unhealthy habits e.g., tobacco use or unwholesome food [55], monitoring abnormal behaviours e.g., fall detection [56, 57] or movement difficulty due to ageing or illnesses [58].

2.3.2 Challenges in Human Behaviour Recognition

A brief summary of the challenges facing researchers in human behaviour recognition and the most used techniques to handle these challenges are presented in this section. Nugent et. al [59] introduced the main challenges of human activities which are listed in Table 2.1. These challenges need to be considered by specific methods and techniques that will overcome them, and some of these techniques will be reported. It is important to get an appropriate method or algorithm that can efficiently recognise human activities behaviour. The major techniques and algorithms used in this field are summarised in some articles such as Liming et. al. in [60]. They divided the activity recognition algorithms and techniques into two major strands. The first one is based on machine learning techniques including supervised and unsupervised learning methods; these methods using probabilistic and statistical reasoning. The second strand of these methods is based on logical modelling in terms of logical theories and representation formalisms. Additionally, Mahmoud in [61] classify the methods into statistical and computational intelligence techniques. Acampora et. al. in [62] categorised the methods into template matching/transductive techniques, generative, and discriminative approaches. Table 2.2 presents a summary of common methods in activity recognition.

The methods represented in Table 2.2 deal with behaviour activity recogni-

Table 2.1: The challenges in human activities.

The activity	**Description**
Recognising parallel activities	Doing more than one activity at the same time, e.g. watching TV and talking to a friend.
Recognising overlapped activities	Activities overlapped to each other, such while a person cooking in the kitchen and the phone rings he/she will stop cooking for a while until he/she finished the call.
Vagueness in activities interpretation	Interpret similar activities in unique way. Like opening refrigerator door, it may have considered as a cooking or cleaning operation.
Multiple occupants	The place occupied by more than a single person.

Table 2.2: The summary of common activity recognition methods.

The method	**The method classification**
Hidden Markov Model (HMM) [63]	Supervised
Bayes Network [64]	Supervised
Support Vector Machine (SVM) [65]	Supervised
Neural Network [66]	Unsupervised/ Supervised
Various Variants of HMM and Bayes Net [67]	Unsupervised
Fuzzy Logic [68]	Unsupervised

tion and prediction. They are classified into supervised and unsupervised techniques. The main idea of using them is to find the dependence and correlations between the sensory data and ultimately identify the behaviour of the participant. Probabilistic models are good techniques in terms of identifying human behaviour. They are capable of representing random variables, data dependency, and its temporal variation [69]. Several researches used probability-based algorithms to build models of human behaviour activities. For instance, the authors in [70, 71, 72] used the Hidden Markov Model (HMM) to model human behaviour in their research and other authors used Bayesian networks to model human behaviour [73, 74]. Alternatively, some computational intelligence techniques are used to represent and model human behaviour in many researches such as the authors in [75, 76], who used support vector machine to recognise and classify the ADL. In other efforts, neural networks are used to recognise human activities in a smart environment [77, 78]. Different types of fuzzy logic have been used in human behaviour recognition [79, 80, 81, 82]

2.4 Synthetic Composite Indicators

The concept of the synthetic composite indicator (SCI) can be viewed as a model of reducing multivariable inputs into a single and meaningful output that can be interpreted by non-specialists. Aggregating multivariable data seems like a simple issue, but it is a very challenging task. However, "the methods for aggregating vast amounts of empirical data remain rather crude" [83]. The aggregated output usually represents the holistic view using a single numerical score and/or an ordinal rank. To achieve the single output value, the index must go through accurate development steps. In general, these steps involve the included or excluded variables, handling missing data, the given weight for each variable etc.

The composite index is generally used to measure the overall progressive changes of the combined elements. Therefore, the following steps should be investigated to understand the process of building the composite index to measure the changes, which are:

- The similarity process of the composite index elements: There are many steps that may be considered to start the measurement of the index. These steps include: making sure that the available variables are appropriate, these variables are enough and well defined to characterise and explain the all elements and suitable to develop a new measurement index to examine the progress change in a certain domain.

- The clustering Process: The clustering can be used based on the comparison of the similarity between different elements that involved in the composite index using different clustering techniques. The clustering process could use distance measurement methods such as Euclidean, Squared Euclidean or City Block etc.

Researchers have developed many techniques to handle the measurement of the changes in different domains. However, in this research, only aggregation techniques are being considered. Therefore, the following sections will present the literature of the main aggregation techniques.

2.4.1 Statistical Aggregation Techniques

The aggregation techniques have two major types that are usually used, which are additive (linear) and multiplicative (geometric or non-linear). The additive techniques use the summation of all weighted indicators and sub-indicators to create a comparison measurement of the studied domain. On the other hand, the multiplicative aggregation is created based on the power of weighted indicators or sub-indicators. Moreover, any additive aggregation technique can offer a compensability between its aggregated indicators. For example, if one indicator has a deficient performance then the indicator that has a significant value can cover the poor one, which may result in creating a biassed composite indicator.

Using the weighted methods to develop a statistical composite indicator is crucial to illustrate the outcomes variable. The choosing of the best strategies for weighting variables is still the focus of researchers. For instance, Babbie in [84] recommends equal weights to become the standard when constructing SCIs. Because it is a simple idea, the author in [85]. supports using this technique to build CSIs. The Equal Weighting (EW) is one of the used techniques that rely on giving all variables equal weight. However, there are some authors do not agree with this approach. Cherchye et al. [86] raised many issues of using EW. For example, they argued that if different weights for each variable cannot be gained, that does not mean to constitute SCI using "fundamentally flawed" method. They also argued that the process of SCI development did not rely on EW. The authors in [87] discussed the use of EW in both approaches: it can be used for perfect exchange among variables and it cannot be used with other variables that will disregard the balancing of these variables.

2.4.2 Computational Intelligence Techniques

Computational Intelligence (CI) techniques are becoming the focus of the researchers because of their precision in predictions, clustering, modelling and trend analysis. A brief review of these techniques related to creating SCIs is presented in this section.

The most popular technique that is used to construct a composite index is the use of Artificial Neural Networks (ANN). For instance, Leigh et al. in [88] presented a work that showed how they used ANN to aggregate multi-variables to form a single output. The authors applied the feed-forward neural network with back-propagation learning into tier data to provide an improvement in the results. The authors in [89] presented an ANN model, trained using Istanbul Stock Exchange (ISE) National 100 Index data to predict the direction of stock price index movement. Such research shows that the use of ANN can model any functional linear and non-linear relationship.

There are some studies using fuzzy logic to forecast data and build composite indexes. For example, Aladaga et. al in [90] presented the Index 100 to measure the changes of the exchange market of Istanbul (IMKB) using fuzzy sets. The authors in [91] introduced a model to predict the future trend of the exchange market in Taiwan using heuristic time-invariant fuzzy rules. A new fuzzy time series forecasting method is proposed in [92]. The authors compared the performance of fuzzy logic with some other techniques and they claim their method gives results that are more accurate.

2.5 Human Behaviour Abnormality Detection

Several studies have investigated methods to recognise normal and abnormal human behaviour activities [93]. The anomaly behaviour can be defined as an unexpected behaviour of the daily routine of a monitored person. The unexpected activities or patterns in the data are often referred to as anomalies or outliers [94]. Many techniques to detect anomaly behaviours have been investigated in many domains including industrial applications, computer networks, medical, image processing, human behaviour etc. Some literature on anomaly detection techniques are reviewed in following sections.

2.5.1 Statistical Techniques

Many statistical techniques are used to model and analyse the daily activities of participants in smart environments. Some of the researchers used statistical techniques to detect abnormality or outlier activities. In the following sections, some statistical techniques used as outlier detection techniques are introduced:

2.5.1.1 Hidden Markov Model

Mori el. [95] applied HMM techniques to model the human behaviour and detect the abnormality in the daily routine of the participant They used likelihood and k-means to describe and segment their data. As a result, the model is able to detect some of the anomalous behaviours in their data. However, the limitation of using their approach is that they did not take into account the differences in the duration time of the behaviour, which could be an unexpected behaviour pattern and their method will consider it as abnormal behaviour. Shin et. al. in [96] developed an HMM model to represent human behaviour and predict unexpected patterns inside a smart home.

Using a similar tree structure, Hierarchical Hidden Markov Model (HHMM) can be used to determine abnormal behaviour activities. For example, it can be used for the duration of the activities and the relationship between them to distinguish between the normal and the abnormal behaviour. Karaman et. al. in [97] established a hierarchical model to monitor people who are suffering from dementia using wearable devices and then they combined automatic motion-based segmentation of the video and activity recognition by a hierarchical two-level HMM to analyse the data. This model consists of multi-model Markov sequence to deal with visual and audio features. They found the combination of video and audio devices to record data from people is very difficult task. This method has many limitations such as the recording of video data can be affected by strong motion and sharp lighting changes which often appear.

HMM can also be used in combination with other methods. For example, Antonakaki et al. [98] used a combination of HMM and SVM to classify human behaviours as normal and abnormal. They proposed to use one class of support vector machine to detect the abnormal behaviour while they used HMM as a classifier for trajectories.

2.5.1.2 Gaussian Mixture Model

Gaussian Mixture Model (GMM) is used in many researches to model human activities that are monitored using sensor devices. For instance, researchers in [99, 100, 101] have used GMM to propose a framework that can recognise motion primitives to create activity models. Both of them used GMM to classify the activities based on wearable sensors. Cardinaux et. al. in [102] used GMM to learn human behaviour and distinguish between normal and abnormal behaviours of an individual person living in a smart home.

2.5.1.3 Other Statistical Models

Several researches have used different statistical methods such as mean and standard deviation to identify abnormalities in their data. For example, Jakkula et. al. in [103] introduced a method to detect the abnormality of using the power in smart environments based on standard deviation. The Allen's interval method in combination with data mining technique is used in [104] to detect anomalies in ADL. They used a small set of data and their results are not yet validated. In general, mean and standard deviation are not good methods to detect the anomalies in human behaviour if the data is not normally distributed.

Another approach of using statistical methods can be found in using the box plot, which is a graphical method that can represent data sets. For example, the box plot is used to display the results of fall detection of an older adult who lives in a smart home [105, 106]. Each box plot summaries data in five principal elements, which are: lowest value, lower quartile, median, upper quartile, and highest value. It will not make an assumption about data distribution when it comes to displaying the differences between populations. However, the box plot is not an appropriate method for every data set.

2.5.2 Computational Intelligence Techniques

A summary of the literature using computational intelligence techniques in abnormality detection is presented below under different section headings.

2.5.2.1 Artificial Neural Networks

Artificial Neural Networks (ANNs) are used to distinguish between normal and abnormal human behaviour patterns from low-level sensors. For example, Mehr et al. in [107] compared three algorithms of ANN to demonstrate the ability of ANNs to recognise human behaviour and distinguish between normal and abnormal behaviours. They applied Quick Propagation, Levenberg Marquardt and Batch Back Propagation on real data sets. Experiments are conducted on real data collected from low-level sensors installed in a smart home occupied by an older adult. They noticed that abnormal behaviour can be detected by comparing the user's behavioural patterns and recent activity that appeared within these patterns. However, their data set is small and more data could change the results because of the behaviour of the ANN, which can adapt its hidden layers to hold additional information if an example is not found to fit the existing structure.

2.5.2.2 Data Mining Techniques

Data mining techniques are used to cluster the data representing the human behaviour and using their association rules to find the relationship between human activities. For instance, Sebastian et. al. in [108] proposed a data mining technique to detect anomalous behaviour in a smart home occupant, with the data extracted from sensors in a real environment. However, their data is not described well. A framework combining statistical analysis with two data-mining techniques is introduced in [109] to identify valid window operational patterns in a smart office and they measured the relationship between human behaviour and energy consumption. Additionally, Fuzzy c-means and k-means clustering techniques are used in [110] to identify the boundaries of normal behaviours after the human behaviour have been recognised using HMM. The research introduces a recognition system based on 3-dimensional data collected by Microsoft Kinect sensor. The data are trained to identify the normal behaviour of the person at his home. The system will consider any new data that does not belong to any cluster as abnormal behaviour.

Zhang et. al. in [111] used different clustering analysis techniques such as hierarchical clustering and adaptive methods that can update the likelihoods of activities based on new observations of action clusters. These techniques are used to compare sensor data sets collected from nine independent living homes for the routine of meal preparations. The data sets represented six month of sensory data. They compared the sensory data against two weeks of ground truth information provided by the residents. The results of the clustering techniques show good potential for using this system on such data. The multi-task clustering approach is used in [112] to classify the everyday activities from visual data gathered from cameras in an intelligent environment. A framework is proposed to tackle the challenging problem of egocentric activity recognition and detect abnormal activities. They presented two algorithms: Earth Movers Distance Multi-Task Clustering and Convex Multi-Task Clustering. They used simulated data to evaluate their methods along with two real data sets. The results achieved the best performance when some a priori knowledge about the data relationship is available. In another effort, Li et. al. in [113] developed a system able to automatically learn the propositional logic rules of intricate relations in complex activities. Their results show that the proposed system can recognise complex activities involving the incomplete or incorrect observations of atomic actions.

2.5.2.3 Support Vector Machine

Support Vector Machine is used in many researches to solve the problem of detecting abnormal human behaviour in data collected from intelligent environments. For instance, Hu et. al. in [114] developed an algorithm to automate the detecting of home visits to an older adult who lives alone in a smart home and report any abnormal visit to this home. They used an ambient and wearable sensing network to collect their data. In addition, they used visiting reports from caregivers to label the positive data. Then they used one class of SVM to train the positive data and tested the algorithm to find the abnormalities in the data. They tested their system in a real-life health care system and it gives satisfactory results when the a system has combination of sensors.

A system for fall detection is introduced in [115]. The system is based on data collected from a home occupied by an older adult person and it consists of two stages. In the first stage, features are extracted from data using ellipse fitting and a projection histogram. The axes of the ellipse are used to distinguish different postures of the human. Then, in the second stage, the extracted features are fed into a directed acyclic graph support vector machine for posture classification of the human. The results from the classification are combined with the derived floor information to detect the fall. They used 15 data sets representing individual people to train the system, then they applied it to a simulated home environment. The results achieved high fall detection.

In another effort, Jakkula et. al. used Support Vector Machine in [116] to detect an abnormality in human behaviour using data collected from an intelligent home. They aimed to develop a system that will help the inhabitants to live in their own home with as little interference as possible. They incorporated Support Vector Machine in Weka Software to achieve better results.

2.5.2.4 Fuzzy Systems

Fuzzy systems are used in many researches to detect abnormalities in human behaviour when using smart environments. For instance, Yao et. al. in [117] used a fuzzy system to build a framework based on fuzzy machine vision to recognise human behaviour. They extracted visual features from the video sequences of the human silhouette, then they used these features as inputs to the fuzzy system. They employed fuzzy c-means clustering to learn the membership functions of their system. They used the Weizman human action data set [118] to test their system. However, their system is not dealing with the high levels of uncertainty in a dynamic real-world environment.

A multi-sensor intelligent system that uses information from several sources (video, audio and other sensors) is proposed by Castro et. al in [119]. They used their system to identify dangerous or interesting intrusion behaviours. Their framework was designed as a generic ontology to receive and integrate as input the heterogeneous knowledge and deal with them as a homogeneous set. They proposed a rule-based model to process the information obtained from the monitored environment. They claim that the model is adjustable to meet the specific needs of each customer and the rules can be configured depending on each scenario. Whenever an intrusion is detected the system will generate an alarm, and it will notify the user via mobile devices. Therefore, the system will generate reports in real time as well as a context-sensitive notification.

In an other effort, Acampora et. al. in [120] proposed a hierarchical architecture framework based on a tracking algorithm, fuzzy inference systems and time-delay neural networks. They aimed to improve the performance of the human behaviour analysis systems in terms of scalability, robustness and effectiveness in behaviour detection. They used the aforementioned methodologies to enable quantitative and qualitative behavioural analysis and for better identifying the activities. They used the well-known CAVIAR data set to validate the proposed framework and compared with other similar approaches working on the same data set.

2.5.3 Distance and Similarity Measures

Similarity measurements are also used to detect anomalous behaviours in intelligent environments [121]. Several researches investigated different techniques and proposed different models that mainly using similarity and distance measurements methods on a wide range of applications such as suspicious behaviour in crowded places, fall detection and abnormal ADLs. However, these techniques still need more investigation when they come to be applied to human behaviour in assisted living environments and specifically human behavioural pattern recognition.

Candás et. al. in [122] proposed a method based on physical activity measurements that automatically mine the data. The method can detect an abnormality in human behaviour by measuring the change of physical activity by comparing the current activity with the historical data. They used real data recorded over two days for twelve people. However, they still need to validate their method by considering activities disorder and the changes of the human behaviour in different days such as the differences between weekdays and weekends. They also need more data to cover long-term monitoring to have a comparable historical data.

In the aspects of sensor application, a similarity function is used to detect abnormal behaviour [123]. The authors model an episode that contains a series of events that includes spatial and temporal information. They used the similarity function to compare two episodes taking into consideration temporal aspects. The proposed method to reduce the wrong classification is to divide episodes into two groups. Weights on individual functions that comprise the similarity function are determined experimentally to ensure the method will get good results. However, the data sets of their work are not clear to fully understand the method and discuss results.

Meng et. al. in [124] proposed a framework to recognise ADL, named the online daily habit modelling and anomaly detection (ODHMAD). In addition, they proposed a habit modelling and anomaly detection for a solitary older adult. Their system works online to identify older adult ADLs and dynamically model their habits. The model will recognise the abnormal behaviour by comparing it with the learnt daily habits. The online activity recognition algorithm is relying on the modelling of the sensors' activation status to determine the occurrence of activities. However, the system needs more investigation and validation.

In another effort, Janetzko et. al [121] introduced an unsupervised anomaly detection algorithm to analyse power consumption data. They also visualised the results of the anomaly scores to guide the analyst to important time points. They used a comparison method to detect the abnormalities in the data. They used different types of visualisation to show the power usage, then combined with a discussion to encode the anomaly values. They tested the method on real data sets collected from a hierarchical sensor network.

2.6 Quality Control-based Signal Drift Correction

Quality control (QC) is considered as an essential step in the large-scale and long-term data processing study, allowing to evaluate the response signal drift, accuracy and retention time during the experiments phases. Based on a controlled experiment involved QC samples throughout the data collection process, the most advantageous uses of QC samples can be obtained, allowing the correction of signal drift and other systematic noise through mathematical algorithms [125].

The signal drift refers to the situation in which the process is changing while being analysed. Techniques used in processing data tend to assume that processes are always in a steady state. But, in real life processes evolve and change over time [126, 127]. Several types of research have applied the concept of drift techniques in different disciplines. For instance, Horányi et. al. in [128] used Observing System Experiments to evaluate the forecast impact of sealevel pressure observations from drifting buoys. They performed experiments of control and denial with and without assimilating drifting buoys' sealevel pressure observations. They found that drifting observations can be used to reduce the forecast error on complex or rapidly evolving cyclogenesis. In another effort Al-Habaibeh et. al. in [129] used infrared imaging data combined with white light image processing to provide an autonomous quality control system as well as a quality assurance system. They found the system can detect the faults in the sealing process.

Edthofer el. al. in [130] reported, the drift is only detected after increase or decrease of the values in a data stream over a period of several days. The statistical significance of the detected drift is then evaluated against the variations in the result and only drift substantially larger than the variations is considered as such.

2.7 Methodology

One of the major challenges in this project is to identify the progressive changes of occupants' activities (or inhabitants) at any point of time. Trend analysis techniques are reasonably reliable techniques for general observation of the changes in activities but not applicable for combining more than one activity to give the observer a holistic overview of the changes in all activities. Therefore, alternative mechanisms are needed for indoor tracking of progressive changes in daily activities. However, the final goal in the investigation of indoor tracking is to invent solutions as efficient as presenting a holistic view of the changes in the daily activities.

For observing the progressive changes in daily activities and to simplify the problem, in this research, trend analysis techniques are used as the first step to look at the changes that may happen in each daily activity. Therefore, the changes that may happen in each activity of different daily activities occurred by the occupants are detected and identified.

The scenario of this research is to use the occupancy where only a single occupant lives or works in the environment. Simple sensory devices including motion detection or Passive Infra-Red (PIR) sensors, temperature sensors, door contact sensors, and light intensity sensors are employed in the absolute single-occupant environment along with a wireless network to collect the occupancy data and environmental attributes in the scenario. An occupancy signal which incorporates both the occupied location (area) and the time of occupancy (with spatio-temporal characteristics) is then shaped and passed to the analysis techniques to identify the progressive changes in the activities and predict the further direction of the trends in the activities.

To extend the single activity analysis to a multiple-activities, combing mechanisms are experimented and a composite indicator is used to combine the multiple activities. In this research, Syntactic Composite Indicator (SCI) is used to identify the changes in the behaviour based on combing all recorded activities on daily bases.

Furthermore, a statistical model of occupancy in the ambient intelligent environment is proposed and implemented to generate occupancy signal for different

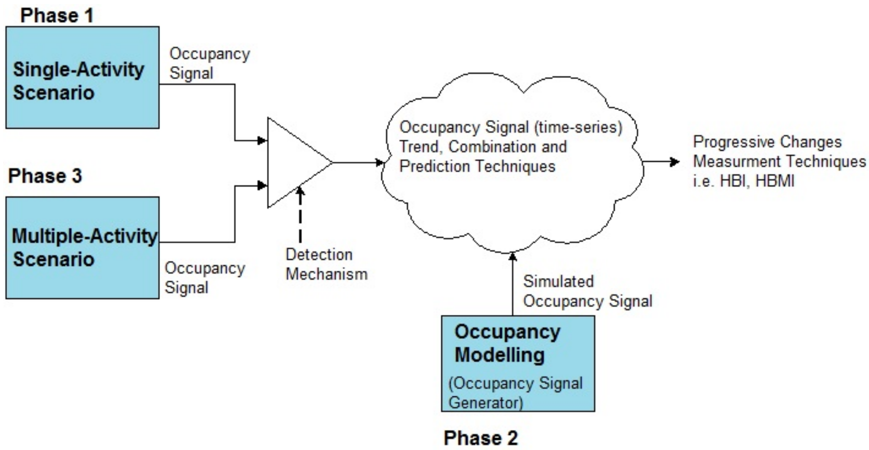

Figure 2.1: Phases of the research (overall picture).

occupant's profile as well as different environmental layouts which can help to train and find best analysis techniques for trend analysis and progressive changes in human activities. In this research, several real data sets collected from real environments are used to prove the performance of the proposed techniques.

Figure 2.1 shows a diagram representing the phases of this research including phase 1: Single-Activity Scenario, phase 2: Occupancy Modelling, and phase 3: Multiple-Activity Scenario.

2.8 Discussion

Based on the knowledge gained from the review in this chapter, it is found that many studies paid attention to the possibility of creating human behaviour models for people occupying an environment equipped with sensors. The literature shows that statistical methods are used in extracting and predicting human behaviour, but the utilisation of these methods could have some issues when it comes to solving complex problems. For instance, HMMs work well to model human behaviour but will not work properly in extracting multiple interacting activities. In addition, more investigation is needed to benefit from similarity measurements to monitor the progressive changes in human behaviour.

Trend analysis techniques are useful in terms of monitoring the holistic overview of human behaviour and to observe the changes in the behaviour that could occur in the long-term of human life. The composite indicators have proven their ability to gather different features of the activities to measure progressive changes in human behaviours.

Computational intelligence techniques are able to address the relationships between the activities and they can be used to detect, recognise, and classify the behavioural patterns of an occupant in a smart environment. An overview of abnormality detection methods is presented in this chapter and the sample of studies are presented here to discuss the usage of these techniques and their limitations. Some of the reviewed studies did not use good size real data sets to validate their results and some of them were carried out and implemented using simulated environment data.

Chapter 3

Data Preparation and Data Representation

3.1 Introduction

In smart environments, it is essential to identify when and where the occupants carried out the most of their activities, in order for them to get more support. Also, identifying their behavioural patterns will help supervisors/carers to determine any future abnormal patterns within their daily routine [131]. For instance, if the person spent a longer time in the bedroom compared to his/her daily routine, or if there is unusual absence for extended periods, then such activity could count as an abnormal activity and the supervisor can act according to it as soon as the activity is observed. Therefore, it is vital to investigate a robust system that can be used to monitor human behaviour based on the daily activities of the people who are occupying smart environments.

The collected data sets from smart environments contain a large amount of complex sensory data that represent ADLs or ADWs of an individual in these environments. It is a challenging task to understand and interpret the human behaviour by extracting them from the low level sensory data. This can be achieved with using one of several techniques or methods that can deal with sensory data.

An individual user behaviour can be modelled based on the sensory data. It is an essential step to represent the data in an appropriate format before any data processing. This process will help to identify the behavioural pattern of each person and it will help to visualise the data more effectively. The data processing will summarise the sensory data in an efficient way for the interpretation and visualisation. Data preparation and data representation are key steps in this research. They will deal with the original data sets before they can be used later in the analysis stages.

In this chapter, different data representation and visualisation techniques are investigated. Without any loss of generality, the discussion presented in this chapter concentrates on the usage of motion sensors and door entry sensors only. This chapter gives an overview of the data pre-processing and data representation and explains how data sets are handled in this research. The description of the procedure of building the simulator that is used to produce synthetic data is also presented in this chapter. In terms of understanding the real data sets that are used in this research, data visualisation in several forms is applied and presented in this chapter. In addition, as a key stage in this research is dealing with and understanding human behaviour trends in the real data sets used in this research, trends are integrated within the data sets provided by the proposed simulator.

This chapter is organised as follows: in Section 3.2, an overview of ambient intelligence is presented. Section 3.3 introduces the data pre-processing and data handling for an intelligent environment. In Section 3.4, a description of sensor data representation approaches is presented. The proposed simulator is presented in Section 3.5 where two different environments are explained in detail. Some conclusions are drawn in Section 4.6.

3.2 Ambient Intelligence Environment

An Ambient Intelligence (AmI) environment is a place equipped with a network of sensors that can be used to monitor the daily activities of the participants in this environment [132]. The AmI can help to improve the quality of people's lives by employing this technology using adaptive intelligent methods to connect between the interconnected embedded systems and services. The major computing areas that are involved in an AmI are ubiquity, intelligence and context awareness. This

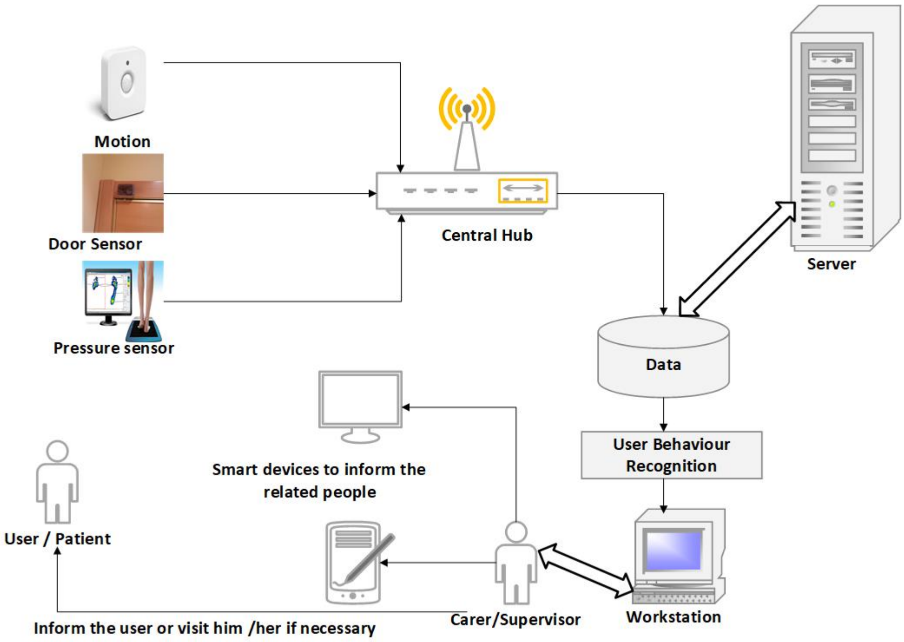

Figure 3.1: An overview of an AmI architecture.

ubiquity is because the participant is using a pervasive infrastructure that has a network of sensors. The intelligence systems could include one or more learning algorithms, pattern or speech recognition, gesture classification and situation assessment. In addition, context awareness may include tracking objects and finding the relationships between objects and their environments [133].

AmI technology can be used to support people taking decisions to benefit the occupants of that environment based on analysing real-time information or historical data recorded from the occupants. Figure 3.1 illustrates a general overview of an AmI environment architecture. This figure shows that the sensor network is used to collect data and communicate with a central database to store the data. Then the system will use the stored data to interpret it into a suitable form and show the results to the responsible person.

3.2.1 Real Home Environment

Tow real homes are used to collect data. Both of case studies are used to validate the results presented in this thesis. In these case studies, each occupant is living

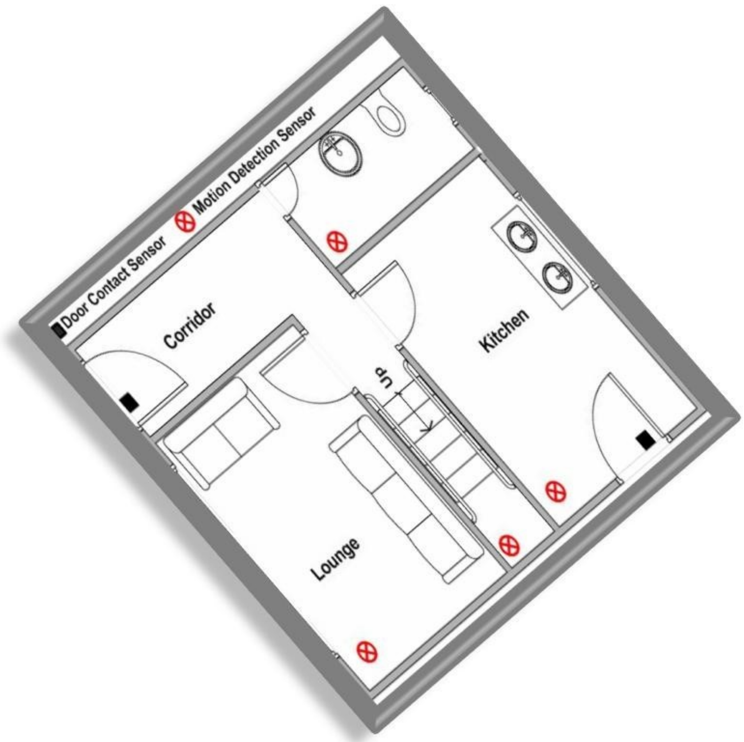

Figure 3.2: Layout of the house and location of sensors of case study I.

alone in different real environments where their movement activities are different from one to another. More detailed description of these case studies is presented below.

3.2.1.1 Case Study I

The data for this case study is collected from a real home equipped with JustChecking monitoring system [134], for an elderly person. The layout of the home environment is shown in Figure 3.2. For this environment, two door entry sensors including front door and back door as well as four motion sensors including kitchen, lounge, upstairs, and bathroom are used. The occupant person was first prescribed some medication, and her health status got worse. Consequently, she was roaming around during the early hours of the day, and her behaviour was considered as abnormal. Then her first medication was replaced by new medication and the patients health got better.

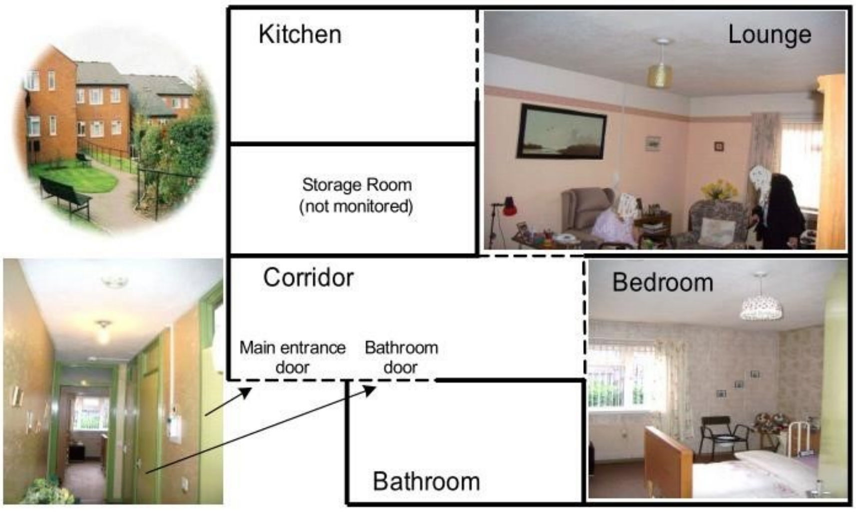

Figure 3.3: Layout of the house and location of sensors of case study II.

3.2.1.2 Case Study II

The data is collected for another real home occupied by an elderly person based on a system developed by [135]. The apartment is in a council complex in Nottingham in the United Kingdom. The elderly person uses a walker support to help her to move around her apartment where most of her ADLs are carried out. The layout of the apartment of this elderly person with some pictures is shown in Figure 3.3. Four motions sensors covering the lounge, kitchen, bedroom and corridor are used. Additionally, two door entry sensors were used to monitor the bathroom and the main entrance doors.

3.2.2 Real Office Environment

The data used in this research are collected from office environment contains four academic staff offices at Computing and Informatics Building, Nottingham Trent University based on a system developed by [136]. Figure 3.4 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 experiment environment. The sensors include a Motion

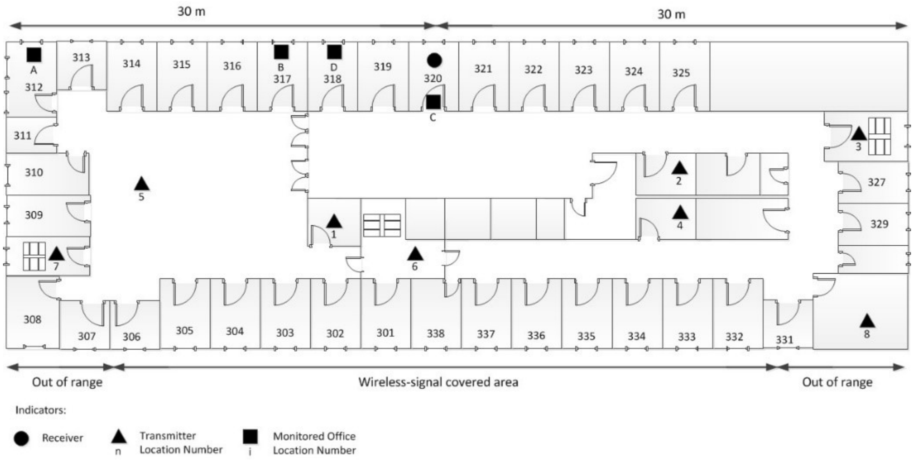

Figure 3.4: Layout of the offices building environment.

sensor, Door sensor, Windows sensor, Pressure sensor measuring chair occupancy, Room temperature, Room humidity, Outside temperature, Room light intensity, Ambient light intensity and Radiator Heater Temperature. These sensors provide information regarding the occupancy and ambient conditions.

3.3 Data Preparation

Data preparation (pre-processing) is amajor step because inadequately controlled methods of data-gathering can result in out-of-range values, impossible data combinations or missing values. Analysing data that is not carefully screened can produce misleading results. Thus, the representation and quality of data comes first before running the analysis. The steps of data preparation can take considerable amount of processing time in each project [137].

The purpose of data preparation is to transform data sets so that their information content is best exposed and formatting to be used by this research's software. Therefore, to prepare data, many aspects should be considered [138], such as:

- Before commencing an analysis, data should be quality controlled. (e.g. via outlier detection),
- For long term data analysis, data series should be as long as possible. Sea-

sonal variability could affect short data series which can give misleading results,

- Missing some data makes the analysis harder and raise questions of data quality. Therefore, handling this issue in the data should be investigated. There are many methods presented to deal with incomplete data series provided that the gaps are not too extensive and that they occur randomly,

- It is important to have frequently recorded data. It could be recorded in hourly, daily, monthly and annual basis. In some cases, the data may be having irregularities.

- Use of summary measures. Summative measures derived from data are often appropriate. For example, calculating monthly means or medians could be very valuable, or deriving monthly maximum values.

3.3.1 Data Formatting

In this research, data is collected to evaluate the performance of the ADL or ADW analysis tools representing a person who lives or works in a smart environment. Data represents sensor activation that show the occupancy behaviour of an adult. The underlying problem in this case, is to monitor older adults in their own homes or adults working in their smart offices using sensors, and it has three main components:

- the human occupant and his/her activities,

- the environment layout within which the person lives,

- the sensors that are used for the monitoring.

The parameters above are proposed to be the key features of representing the participant's behaviour and the environment's configuration. These parameters should be translated into numerical values. For example, the number of daily times of visiting the bathroom; the mean time of occupying the living room. The key parameters should be used in the mathematical model that show these features. The parameters (together with human and environment constants and variables) give the overall state describing the current situation.

Table 3.1 shows the formatting of data summarisation, which is representing activates of an older person based on his/her sequential movements and the duration of staying in a specific location. Each behaviour can be represented using two order levels. The first level is the activity sequence that will show the older adult's movements. The second level is the sequence of actions, which is showing the occupancy time of the activity which the older person has performed. For example, walking to the kitchen is a moving activity and sitting in the living room is an occupancy activity.

Activities are presented into five main categories. They are "Bedroom", "Living Room", "Kitchen", "Bathroom" and "Outside" activities representing sleeping, socialising, eating, cleaning and going out respectively.

Table 3.2 shows the formatting of data summarisation, which is representing activities and their durations for a person who uses a smart office. From the sample of the data set, it is clear that the data gathered from sensors in a smart office sometimes overlap. For example, the chair pressure sensor could be triggered at the same time as the PIR sensor and the PC.

Table 3.1: A sample of real data collected from a smart home including missing data.

Smart Home		
Date and Time	**Sensor Type**	**Location**
$Day1$, 09 : 04 : 17.656	PIR	Living Room
$Day1$, 09 : 24 : 17.656	PIR	Master Bedroom
$Day1$, 11 : 06 : 03.796	PIR	Kitchen
$Day2$, 10 : 04 : 17.656	PIR	Hall
$Day2$, 10 : 14 : 10.050	PIR	Living Room
$Day3$, 11 : 45 : 10.766	PIR	Master Bedroom
$Day4$, 12 : 06 : 03.996	PIR	Kitchen

3.3.2 Handling Missing Data

Data could be missing because of hardware or software faults. Statisticians often use the terms "missing at random" and "not missing at random" to represent different scenarios. Data missing at random (MAR) means that they are missing unrelated to actual values of the missing data. For instance, suppose one of the hardware devices is out of action for some time, this would be unlikely to be related to the participants performing his/her activities.

Data "not missing at random" (NMAR) means that the missing data is related to the actual activity that leads to unrecorded data. For instance, if a participant leaves his/her home for some time, then missing outcome data will be more likely. Such data are "non-ignorable" in this case, as the analysis of the available data alone will typically be biased [139].

To start dealing with missing data the following steps should be considered:

- Use what is known about the data
- Try to find out why data is missing
- Understand the distribution of missing data
- Decide on the best analysis strategy to yield the least biased estimates

There are many statistical methods to deal with missing and outlier data. The common methods include Listwise Deletion, Pairwise Deletion, Mean/Mode

Table 3.2: A sample of real data collected from a smart office including missing data.

Smart Office		
Date and Time	**Sensor Type**	**Location**
$Day1$, $09:04:17.656$	Pressure	Chair
$Day1$, $09:05:03.766$	PIR	PC
$Day2$, $10:01:03.65$	Door	In
$Day2$, $10:02:01.75$	Pressure	Chair
$Day2$, $10:02:01.75$	PIR	PC
$Day3$, $09:05:03.766$	PIR	PC
$Day3$, $10:30:03.796$	Door	Out

Substitution, Dummy variable adjustment, Regression Imputation, Maximum Likelihood and Multiple Imputation [140, 141].

Data examination and exploration before performing and interpreting analysis are essential; therefore, considering ways to evaluate and understand missing data is the first step that should be taken after formatting the data. The following steps are the principal options for analysing missing data:

- Analyse only available data (i.e. ignoring the missing data),
- Imputing the missing data by replacing values, and treat these as if they were observed (e.g. imputing the mean, imputing based on predicted values using analysis methods),
- Imputing the missing data and take into account the fact that these were imputed with uncertainty (e.g. multiple imputations, simple imputation methods with adjustment to the standard error),
- Make an assumption by using statistical models to process the missing data based on their relationships with the available data.

In this research, the Pairwise deletion method is used. It is found to be more suitable to process our data. It is used to improve the validity of these research results and to reduce the waste of resources caused by missing data. The basic idea for this method is to analyse all cases in which the variables of interest are present and it attempts to minimise the loss that occurs in listwise deletion. This method keeps as many cases as possible for each analysis and uses all possible information with each analysis. It cannot compare analyses because the sample is different each time [140].

Table 3.3 shows samples of missing data in our data sets. We will not discard the whole data, instead only the missed value was discarded from analysing the dataset.

3.4 Data Representation

To deal with sensory data collected from an AmI environment and find a method that can be useful and flexible to represent a large amount of data sets generated

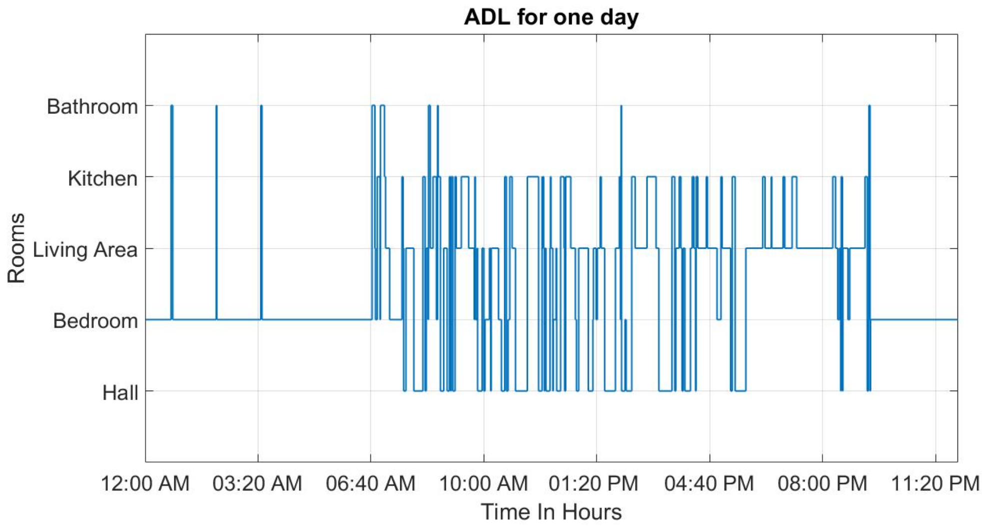

Figure 3.5: A sample of data representation of one day of ADL.

from sensors is one of the main challenging tasks. Also, many other important tasks can be performed using the representation of data sets such as clustering, classification, visualisation, prediction, forecasting and identifying normal and abnormal patterns. Moreover, data sets extracted from a network of multiple sensors are often difficult to understand if they are not processed.

In this research, different data sets from different smart environments are used. Each data set represents an individual participant in one smart environment. The first type of the data sets represents one person living in a smart home and

Table 3.3: A sample of real data with the handling missing data.

Date and Time	**Sensor Type**	**Location**	**Sensor Value**
$Day1$, 11 : 04 : 17.656	PIR	Living Room	00011
Missing Data	PIR	Master Bedroom	00011
$Day1$, 11 : 06 : 03.796	PIR	Kitchen	00011
$Day2$, 10 : 04 : 17.656	**Missing Data**	Living Room	00011
$Day2$, 10 : 14 : 10.050	PIR	Living Room	00011
$Day2$, 11 : 45 : 10.766	PIR	Master Bedroom	00011
$Day2$, 12 : 06 : 03.996	PIR	Kitchen	00011

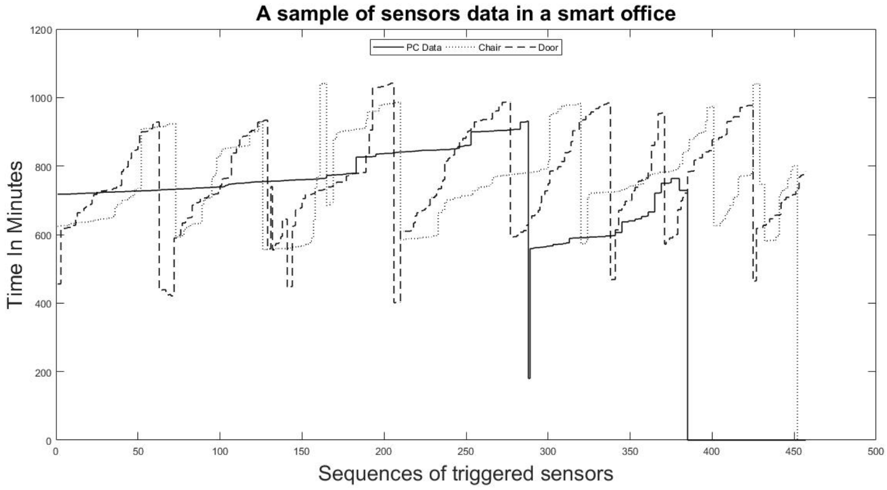

Figure 3.6: A sample of data representation of one day of ADW.

there is no parallel activity in different areas to be detected. The second type of data sets represents a single person working in a smart office and data can be overlapped. The data sets used in this research comprise 120 days of monitoring a person in a smart office and 140 days of monitoring a person in a smart home. Figure 3.5 illustrates the signals from one day gathered from multiple sensors used to monitor a person who is occupying a smart home. Figure 3.6 shows one day of signals collected using a network of multiple sensors monitoring an individual person working in a smart office. In addition, data sets generated by the simulator can have any number of days with different personalities that can simulate a person living in a smart home.

These data sets are used to identify the behavioural patterns of each user and then they are used to distinguish between the normal and the abnormal behaviour based on the correlation between events and activities. Furthermore, trend analysis techniques are used to understand the human behaviour and to measure progressive changes based on the acceptable amount of data sets available for this research. In addition, because of the complexity of sensory data especially when it is collected using multiple sensors, the difficult challenge of interpretation of copious amounts of sensory data could occur.

The transition between different areas in a smart environment (e.g a smart home) will be represented by signals of sequential movements through the en-

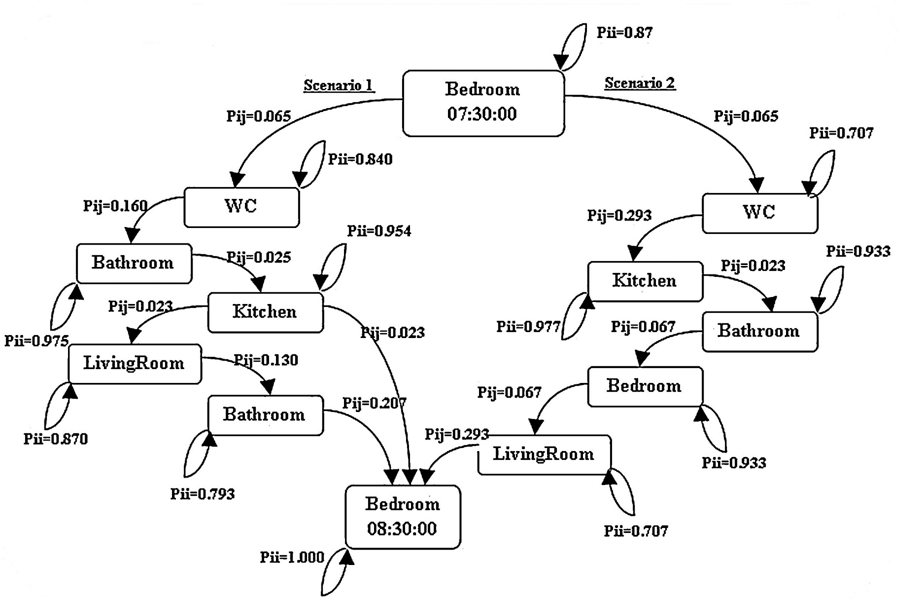

Figure 3.7: HMM sample of movements.

vironment's areas. Signals from a single-occupant occupancy are dependent on the profile of the environment (design of the environment). The occupant may spend some time in each area which is called the duration. A Hidden Markov Model (HMM) is used to represent signals and duration of human activities in this research, because of its ability to mimic human behaviours. HMM is a generative probabilistic function of Markov chains based on the first order Markov assumption of transition. HMM consists of a hidden variable and an observable variable at each time step. The basic idea of HMM is based on the Markov first order assumptions, which is that the future state depends only on the current state [142].

To model an older adult's behaviour using HMM, suppose that the occupancy state is tracked. In this model, the hidden state is discrete and consists of two possible outcomes: "occupied" and "not occupied" for each defined area. It is assumed that the occupancy state of the person can be changed over time and is sometimes unpredictable (as in real life). Assume that if a change of the occupancy happened with probability of (a) then state with probability of $(1 - a)$ remains the same. For example, when $a = 0.1$, it is expected that on average 1 out of 10 of the transitions will be associated with a change in occupancy state. The Figure 3.7 illustrates a sample of the proposed HMM to model the movements of one person.

In addition, a prior $P(X_0)$ needs to specify over states at the first-time step, which is assumed to be uniform. To complete the model, the observations are assumed to be discrete and consist of only two possible outcomes: "duration" and " vacancy" for each area. As might be suspected, the model is parametrised such that duration is more likely in the occupied state and vacancy is more likely in not occupied state. Specifically, with probability of (b) a person produces an outcome that is consistent with their occupancy state is assumed. The model is parametrised such that "occupied" is more likely in a duration state and "not occupied" is more likely in a vacancy state. Specifically, it will be assumed that with probability b, a person produces an outcome that is consistent with his/her an occupancy state. Instead of using symbols, the states "occupied" and "duration" are represented by the number 2 and "not occupied" and "vacancy" are represented by the number 1. Therefore, $X_t = 2$ represents a occupancy state at time t and $X_{t-1} = 1$ represents an observed vacancy at time t. Equations 3.1, 3.2, 3.3 and 3.4 are representing the human activities model using HMM [143]:
State Probability

$$P(X_{(n+1)} = j|X_n = i) = P(X_1 = j|X_0 = i) \tag{3.1}$$

and the Transition Matrix:

$$T(i, j) = P(X_t = j|X_{t-1} = i), 1 < t \leq N \tag{3.2}$$

with

$$T(i, j) \geq 0, \forall (i, j) \in N^2 \tag{3.3}$$

and

$$\sum_{N}^{j=1} T(i, j) = 1, \forall i \in N \tag{3.4}$$

3.5 Simulated Environment

The simulated environment is used to generate data sets similar to data sets collected from the real environment without hardware costs. The real environment that is equipped with a large number of tools such as sensors, actuators and computing components to collect data sets is expensive. Therefore, many researchers, working with a simulated environment because of the expensive of real environment tools or because of collecting data from sensors is one of the main steps in self-adaptive applications and sometimes researchers require large data samples to test and justify better techniques for their research which are in most cases will not be found from real environment [144, 145].

3.5.1 Simulation Methodology

The simulator's main objective is to generate data sets that can be used for evaluating the performance of the ADL analysis tools representing an older adult living in a smart home. It must be able to mimic sensor activation that shows the occupancy behaviour. The underlying problem, in this case, is to simulate monitor the person in his/her own smart home as monitored using sensors, and it has three main components as mentioned earlier in Section 3.3.1.

The parameters of the proposed simulator are the key features which represent the older adult's behaviour and the environment's configuration. These parameters should be translated into numerical values. For example, the number of daily times of visiting the bathroom; the mean time of occupying the living room. The key parameters should be used in the mathematical model that show these features. In the simulator, some of these factors could be changed when the defined environment needs to be changed to simulate another layout.

The proposed simulator follows a two-level mechanism where the first level is related to profiling the activities based on the movements of the older adult, and the second level is related to modelling the localisation and movement of each profile. Therefore, the simulator design involves two phases. The first phase is simulating the sequential movements of the older adult and the duration of staying in a specific location; the second phase is to ensure that the simulated data is representative of the person's behaviour. Each behaviour can be represented

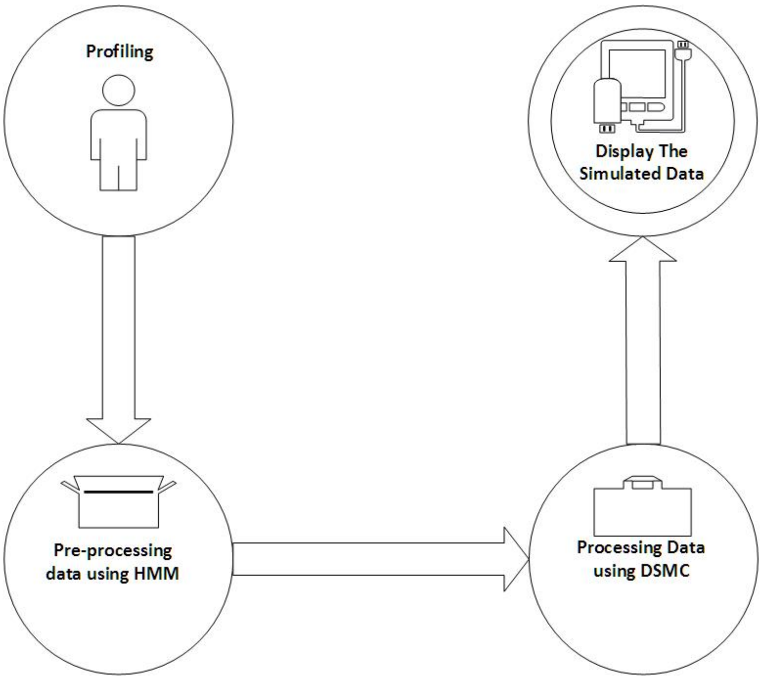

Figure 3.8: Diagram of the simulation stages.

in two order levels. The first is the activity sequence that will show the older adult's movements. The second level is the sequence of actions, which is showing the occupancy time of the activity which the older person has performed. For example, walking to the kitchen is a moving activity and sitting in the living room is an occupancy activity.

The methodology behind the creation of this simulation contains four stages as shown in Figure 3.8. These stages are: profiling, data pre-processing for Hidden Markov Model (HMM), processing data using Direct Simulation Monte Carlo (DSMC) and representing the simulated data. More detail about each stage is given in the following sections.

Activities are presented into five main categories. They are "Bedroom", "Living Room", "Kitchen", "Bathroom" and "Outside" activities representing sleeping, socialising, eating, cleaning and going out respectively. Each of these activities are represented as a state and they are illustrated in Figure 3.9. The transition (link between each state) is also presented in the figure.

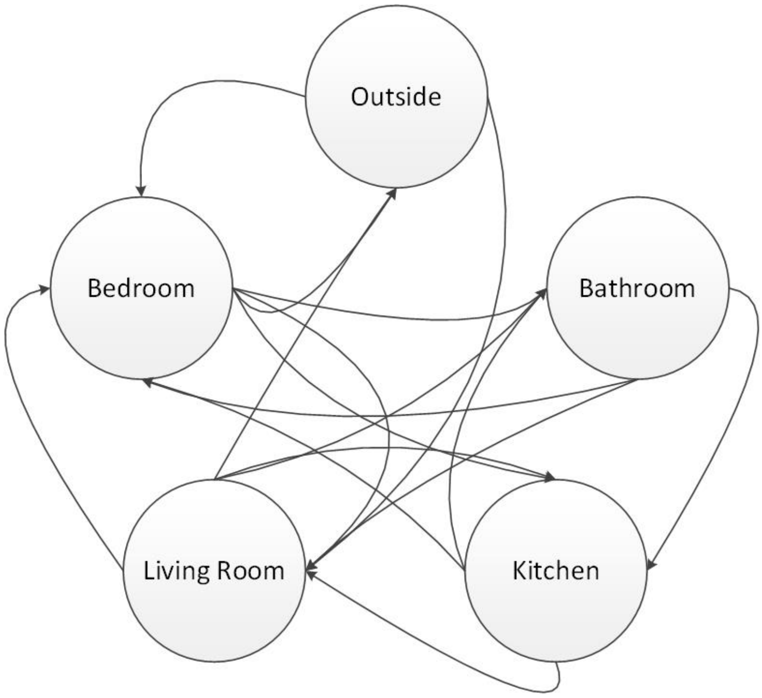

Figure 3.9: State diagram of activities.

3.5.2 Profiling

To simulate different people profiles with different behaviours, a set of parameters describing the characteristics of each person is needed. For example, it is assumed that the duration of staying in the bedroom and the start time of going to the bedroom are different from one person to another. The sequence of visiting different areas is also different for each person. In addition to the the difference in the sequence, the duration of each activity could also be different. Therefore, it is important to identify some parameters in the simulator to represent the user profile. The parameters (together with human and environment constants and variables) give the overall state describing the current situation, and they control the next action to be taken.

The proposed simulation environment is based on a real flat (apartment) environment, which has a basic monitoring system installed. For each activity, the sequence of movement and mean duration of staying in a specific area are recorded. The sequence and duration are stored in a 2- tuple as illustrated below:

$$
\begin{aligned}
Seq \;=\; & [(Bed\,Room, 25); (Kitchen, 2); (Living\,Room, 50); \\
& (BedRoom, 120); (LivingRoom, 60); \ldots]
\end{aligned}
$$

The algorithm will keep the track of the sequential movements and the mean duration of the person's localisation in each area for the whole day. The following aspects are also taken into our account in the simulation:

- The times of movements between the areas for each day is different, and the mean duration time of occupancy is different.
- The algorithm assumed that the occupant has sequential movements starting from one area to another depending on the layout of the environment for a normal pattern.
- Each daily occupancy signal can affect the occupancy signal for the next day. For instance, the next day activity cannot start in bed area if the previous day activity ends in another area.
- Daily observed occupancy pattern can become longer than an expected occupancy pattern, because of interrupts that may happen.
- The first area met on the first day of activity simulation would be the first area of the expected occupancy pattern in the occupant's profile. In the case of unexpected transitions, if they happen many times it may indicate abnormal activity.

3.5.2.1 Direct Simulation Monte Carlo

In this step, the concept of Direct Simulation Monte Carlo [146, 147] is applied on the state-space to ensure that the generated data is modelled as a dynamic system and data covers the entire day. This technique is used to model the evolution of the created system over time, and measurements are assumed to be available at discrete times. For dynamic state estimation, the discrete-time approach is widespread and convenient.

The direct simulation algorithm has two steps - the first step is the sampling proposal for the current hidden state and the second step is to assess the likelihood of the data given each proposal. However, any proposal that may produce unlikely projected distributions for the observed data may be ignored. At the same time, the "prior" distribution from the state transitions and the "evidence" from the observed data are weighted.

This procedure will product simulated data, which is displayed in a matrix of [Duration, Area, Date-Time], therefore the state of each area in the specific time will be known. The following equations were used to consider the occupancy date and time for the simulation:

$$N = S * (j, i) \tag{3.5}$$

where S is the state of each Area

$$j \in [Startday, Endday] \tag{3.6}$$

$$i \in [1, 86400]\, Seconds \tag{3.7}$$

$$N \in [1, No.\ of\ \ Areas] \tag{3.8}$$

3.5.2.2 The Proposed Algorithm

In this work the HMM with DSMC is combined to build a simulator for a single occupancy model to represent the older person's movements and occupancy duration in the proposed environment. This system can provide long-term pattern data, which is needed for our research. Generally, focusing on finding out a long-term patterns and trends in older person's activities are key areas of concern; they are necessary to estimate the progress or deterioration in these activities which could help the medical assistant or caregiver. Here is the summary of the proposed procedure that is used to generate the data:

Algorithm 1 The Simulation Algorithm.

procedure HMMDS
    Input the Start day and End day of the simulator
    Generate the person's sequential movements
    calculate the mean time of each occupancy state
    repeat
        For time t, session m
        repeat
            Sample a random particle i
            $i \sim Uniform(1, ..., M)$
            Sample a proposal $X^* \sim P(X_t|X^i_{t-1})$
            Sample $u \sim Uniform(0, 1)$
            if $u < P(Y_t|X^*)$ **then**
                accept the proposal.
                Set $X^m_t = X^*$
        until all required sessions covered
    until all required time covered

3.5.3 Create the Profiles

As it was explained in 3.5.2, the idea behind it is to create different people profiles and generate different patterns of data. This stage starts with defining the sequential movements between the areas and generating new movements after training with the real data. In this stage, HMM is used.

3.5.4 Generate the simulated data

In this phase, using a predefined function in the HMM toolbox, the primary data is simulated. Then the primary data is handled by the concept of the Direct Simulation Monte Carlo to generate the simulated data in its final format, and store it in a matrix containing the duration, the area and the start time of occupying this area. A sample of the simulated data is presented in Table3.4.

3.6 Validation

The validation step is the most challenging phase; it is not easy to validate simulation models of human behaviour. This process is to evaluate the software during or at the end of its development to determine whether it meet the specified requirements. The author in [148] provides a survey on how to validate a simulation model based on statistical techniques. He assumed that "the type of technique actually applied depends on the availability of data on the real system. Regarding this data availability, three situations should be distinguished: (a) no real-life data are available, (b) there is only data on the real output (not the corresponding input or scenario), (c) besides the output data, the corresponding input or trace is also known, which is used to perform so-called trace driven or correlated inspection simulation".

In this work, the main goal is to create models that aim to provide better understanding of the activities of daily living for an older person. To assess the quality of the model, the proposed approach to validate the simulation statistically is a goodness-of-fit measure. This technique is used to match the proposed model outputs to the real database that is available to the authors.

3.6.1 Goodness-of-fit Measures

There are several goodness-of-fit measures that could be used to evaluate the overall performance of simulation models. Popular among them are the root-mean-

Table 3.4: A sample of simulated data.

Duration in Minutes	Location	Date and Time	Sensor Type
12.229	Bathroom	$2014-2-17:56:15.45$	PIR
66.145	Bed Area	$2014-2-18:8:29.321$	PIR
3.045	Bathroom	$2014-2-19:14:38.674$	PIR
24.964	Kitchen	$2014-2-19:17:40.239$	PIR

square error (RMSE), goodness-of-fit for a Binomial distribution and goodness-of-fit for a Poisson distribution. In this research, the goodness-of-fit for a Poisson distribution is used.

Before this method was used, a day was divided into four time periods; morning, noon, evening and night. Suppose that the mean time and the number of times a particular event occurs in each of the major periods (one of four) are observed. Therefore, the mean time of occurrences of each activity at the period of time could be deduced. To calculate Poisson probabilities the Equation 3.9 was used. Then, the X^2 test was used to see how closely the observed data agree with the simulated data; Equation 3.10 was used to calculate the X^2 test.

$$P(X) = \frac{e^{-m} m^x}{x} \tag{3.9}$$

for x=0,1,2,3,...

Here P(X) means the probability of the activity (x) to randomly occur per time, m the mean time of each activity in the time.

$$cal\ X^2 = \sum \frac{(R-S)^2}{R} \tag{3.10}$$

where R is the real data and S is the simulated data.

3.7 Visualisation of Sensor Data

Data visualisation is used to help in understanding the real data sets as an initial stage of analysing the data. Also, data visualisation shows the possibility of understanding the pattern in these data sets. The use of data visualisation clearly can help to ease the examination of enormous amounts of data and allows conclusions to be made from the relationships within the data. In the next sections, some samples of the techniques used to visualise the sensor data are presented.

3.7.1 Visualisation Using Some Data Features

It is one of the useful visualisation methods which can help to understand the sequences of binary data. Using start-time and duration that are extracted from

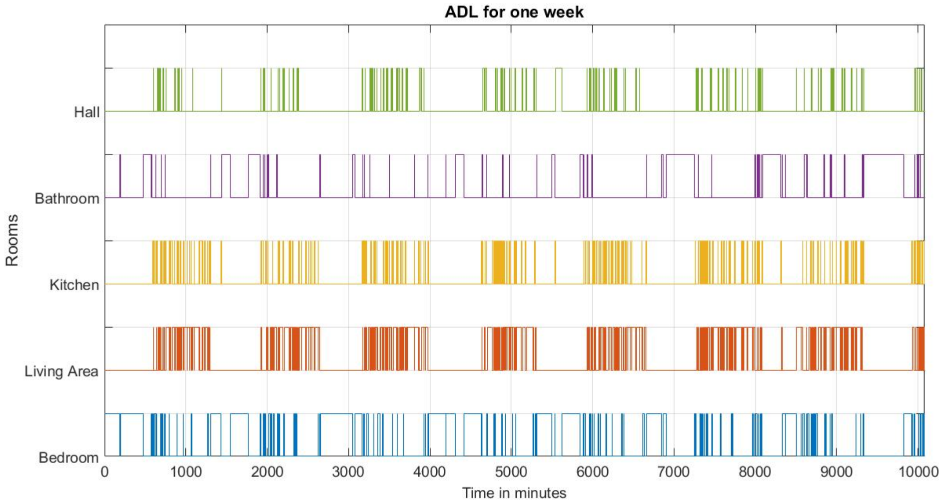

Figure 3.10: Sequences of movements in one week of a person in a smart home.

occupancy sensors one can interpret the sequences of the movements of the person who is using a smart environment and it will show the activities pattern of this person. It is sometimes difficult to understand human behaviour when a big data set is visualised. For example, Figure 3.10 shows one week of duration and movements of a person living in a smart home. It is clear from this example that if the visualisation and tracking of the movements of this person represented more days (e.g. a month) the graph would be very difficult to understand.

If the graph shows one day, for example, it will give a better idea about the behaviour in that day but it will not give a holistic view of the monitored behaviour. For instance, consider Figure 3.11 which shows the movements and duration of one day of monitoring a person in a smart home and Figure 3.12 which shows the plot of a sample data set representing one day of occupancy and duration of the same person. The data sets presented in these figures are collected using five sensors representing the occupancy in a smart home. The behaviour of the user can be more easily interpreted in Figures 3.11, 3.12 than Figure 3.10. For instance, Figure 3.11 shows the sequences of the movements of this person for one day and the duration each time he/she visited any place in the home. It is very difficult to achieve this level of understanding from the raw sensory data displayed in Figure 3.10.

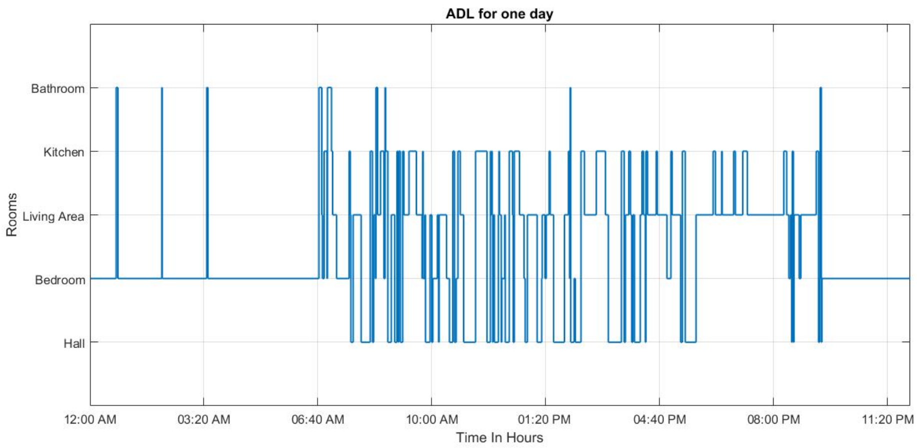

Figure 3.11: Sequences of movements in one day of a person in a smart home.

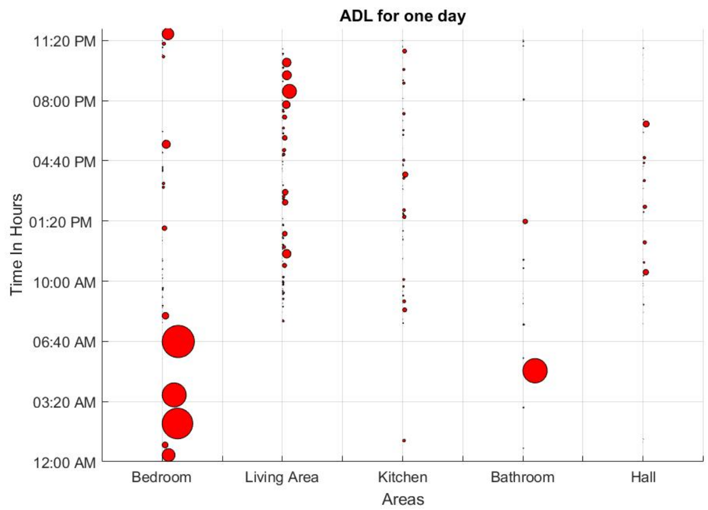

Figure 3.12: Start Time and Duration of one day in a smart home.

3.7.2 Visualisation of The Trends

To have a better understanding of the big data sets that represent human behaviour, trend analysis techniques are used to find the trend within the studied data sets. For example, moving average techniques are used to smooth the data and to find trends in the data. Also, the moving average can be used as a simple prediction method. It is used in this research for different data sets. Figure 3.13 shows how it can identify a trend of the start-time and duration of the data set

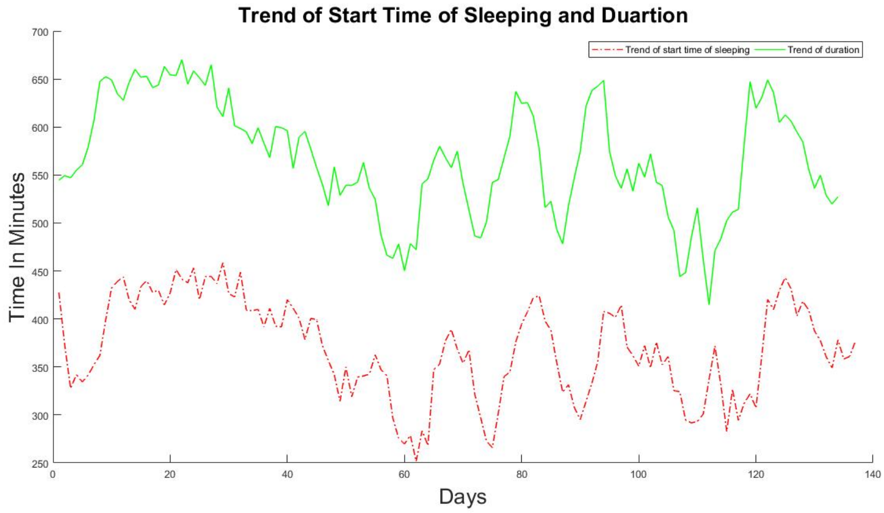

Figure 3.13: The trend of start time of sleeping and the duration.

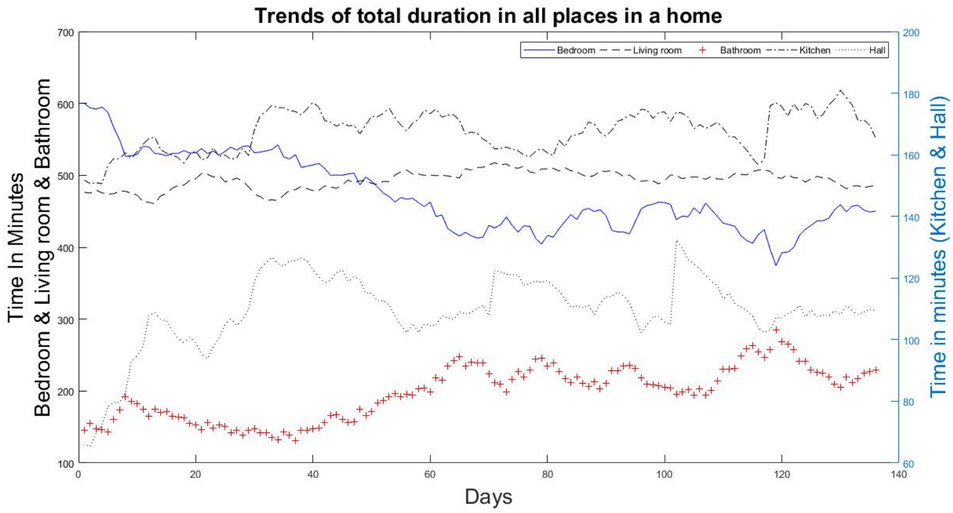

Figure 3.14: The trends of the daily duration start time of sleeping and the duration.

and it is representing the pattern of the monitored person. Figure 3.14 shows the trends of the daily duration of occupancy in each place in a smart home. This kind of information gained from plotting trends can help to make a decision to prevent the worst situation that may happen to the monitored person.

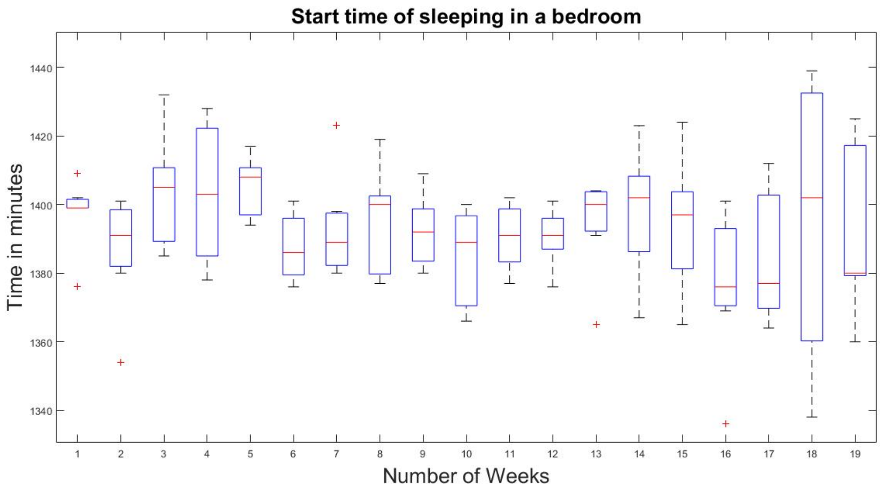

Figure 3.15: The trend of start time of sleeping and the duration.

3.7.3 The Box Plot

The box plot is a method that can be used in data analysis to represent some statistical information on data sets graphically and to identify patterns that may be hidden in a data set. It is used to summarise and compare groups of data in a data set. The box plot uses statistical methods to be displayed in its graphs such as the median, the approximate quartiles, and the lowest and highest data points. It uses these methods to connect the level and symmetry of the distribution of data values [149]. The box plot is used in this research to have an initial understanding of the data sets. For instance, Figure 3.15 illustrates the summary of the weekly start time of the sleeping in a bedroom and Figure 3.16 shows the total duration of the sleeping on weekly basis.

3.8 Discussion

In this chapter, the methods of data pre-processing and data representation are discussed. These methods are primarily employed for human activities monitoring to help them live or work independently in their own environments. The challenging tasks of processing the big data sets collected from a network of sensors are also explained. Some approaches are used to convert binary data sensory

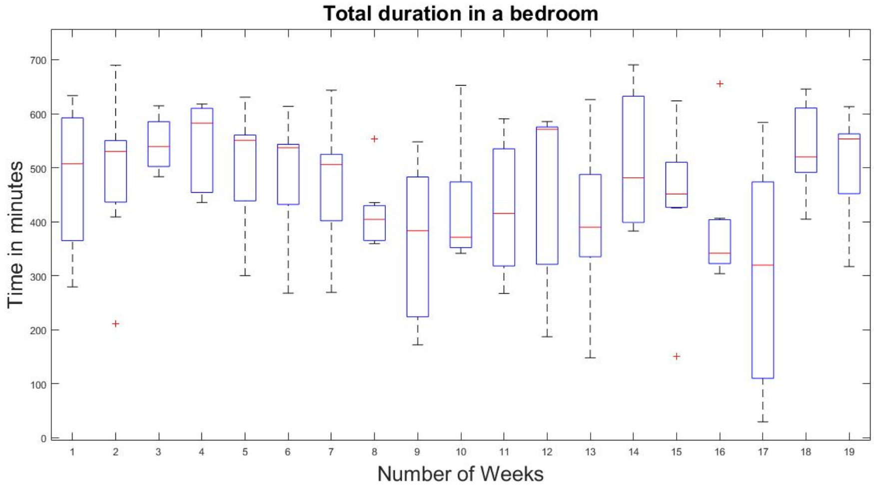

Figure 3.16: The trend of start time of sleeping and the duration.

into floating point values without losing key features after conversion.

Real and simulated data sets are described in this chapter. Some data sets examples from different real environments are presented. A simulator is built to support this research by producing simulated data sets. In addition, the simulator is used to include linear trends on the output signal, which emulates the actual data generated from real environments. In Chapter 5 and 6, the data sets are used to build a system that can measure the evolution of human behaviour.

Chapter 4

Trend Analysis and Prediction Techniques

4.1 Introduction

In this research different trend analysis techniques are applied to data sets that are recorded from monitoring a person who lives or works in a smart environment. Applied trend analysis techniques in this work are described briefly in this chapter.

Trend analysis techniques are developed based on regression and hypothesis testing for long-term recorded data. The time in such recorded data is usually the main explanatory variable of interest for trend analysis through spatial or directional trends. Trend analysis is used to detect both sudden and gradual trends over time. The general purpose of trend testing is to determine if a series of observations of a variable collected over time has increased or decreased in their values (getting "better" or "worse"). Statistically, this is to determine whether the probability of the variables' distribution from which they arise has changed over time or not. It is also to describe the amount or rate of that change. Changes could be in the distribution value such as a mean or median.

Trend analysis is a subject of interest in different branches of science. This includes environmental study, health care studies, economic analysis, and it is rarely investigated for human behaviour in terms of monitoring ADLs or ADWs in smart environments. Trend analysis could also be used to analyse humans' general health surveillance, monitoring, activity evaluation and well-being. For example, analysing activities of an older person can only be fully understood if their daily activities are examined in terms of person, place, and time. A study analysing human behaviour trends in activities of daily living may focus on one or more of the following:

- The overall progressive change of behavioural pattern - the general goal of trend analysis for an older person's well-being health surveillance is to distinguish the status change over time and if it has changed, how quickly or slowly the changes have occurred.
- Comparing one period to another - this type of trend analysis is used to assess the health of an older person before and after a specific period of time. Evaluating the impact and status of the person by using medical or other technical devices, which is sometimes called interrupted time series analysis.
- Comparing one geographic area to another - when comparing occupancy and usage of an area in a smart environment, looking at one point in time can be misleading. Analysing the trend over several months can give a more precise comparison between areas. For instance, one area may have a higher value on an indicator in one month, but a lower value in the next month.

Based on existing techniques that are used in other branches of science (e.g. financial trading) such as moving average, seasonal test and momentum methods, new techniques are proposed to deal with the challenges of this research. The investigated techniques are applied to data sets that collected from monitoring a person who lives in a smart home or who works in a smart office.

4.2 Time Series

To analyse a signal or time series, it is essential to identify its characteristics. The identification of the characteristics will help to determine the trends within the

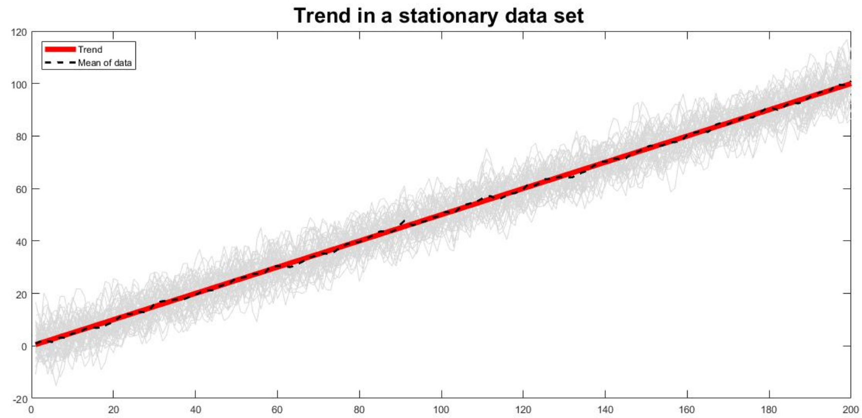

Figure 4.1: An example of a stationary time series with a trend.

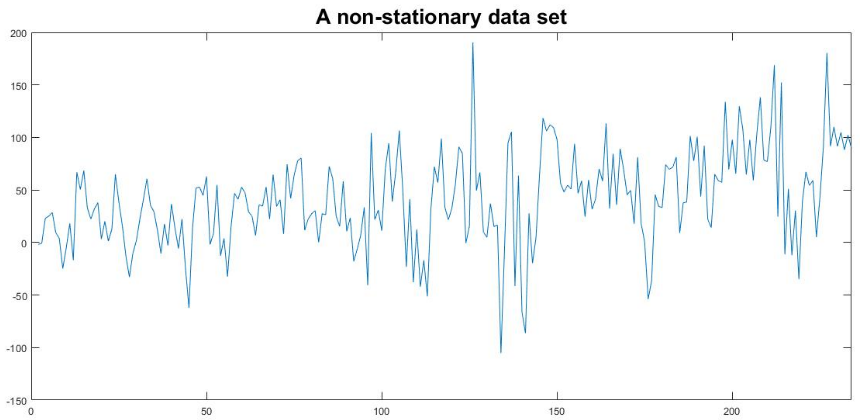

Figure 4.2: An example of an non-stationary time series.

time series. To determine the variability in time series, the phenomenon that acts as the essential cause of the general tendency and non-essential random causes should be analysed, which determine the degree of deviation from trend. Variability can identify the type of series (stationary or non-stationary). In addition, If the series is non-stationary, the first step is to select the most appropriate model to determine the trend that will help to deal with the time series as stationary, and that helps to choose the most appropriate model to analyse and forecast the time series. For a signal or time series to be stationary, the expected value must be a constant and the co-variant is just a function of delay (lag) and not time. For non-stationary signal and time series, it is essential to separate different elements [150]. In general, any time series, $Y(t)$ could be represented as a combination of the elements expressed below 4.1;

$$Y(t) = trend + seasonality + random\ fluctuation. \tag{4.1}$$

An example of a stationary time series with a linear trend is presented in Figure 4.1 and a non-stationary time series is presented in Figure 4.2.

4.3 Trend Analysis Techniques

The trend represents a gradual change in the pattern of data; this includes its condition, output, process, an average or general tendency to move in a certain direction over time. Trends could be represented by a line or curve on a graph [151]. The authors in [152] defined trend as "trend is long-term temporal variation in the statistical properties of a process, where 'long-term' depends on the application".

Many monitoring programs have an objective to detect changes or trends in levels of what is being monitored over time. The purpose may be to look for increased or decreased values resulting from changes in specific objectives' behaviour. The empirical approach of trend analysis is to look for relationships that could explain how the system works or to test hypotheses suggested by process-based considerations. In particular, the trend analysis is the preparation of collected information and attempting to identify a pattern, or trend, in the information.

Predicted future events are often obtained by trend analysis. It could also be used to estimate uncertain events in the past, such as how many times the monitored person visited a specific area (e.g. toilet) during the sleeping period on different days.

The common types of trends are summarised in [138], which are: 1) a sequence of measurements with no trend, 2) fluctuations (moved up and down) along the sequence, 3) cyclical pattern trend, 4) linear trend (could be rising from fluctuations cycle). The following sections will give an explanation of the techniques used in this research.

4.3.1 Moving Average

In statistics, Moving Average (MA) is a technique used to get an overall idea about the trends in a data set and it is useful for forecasting long-term trends [153]. The MA creates a series of averages of different subsets of the full data set and it can be calculated for any period. For example, if data is available for years of monitoring an older person in his/her smart home, then a moving average of days or months or years could be calculated. Using long-term monitoring data to calculate moving average will help to see trends in the user's behaviour and forecast his/her well-being situation.

To calculate the MA, an average of a fixed subset size from a series of numbers should be calculated, and then the subset is modified "shifting forward" by excluding the first number of the series and including the next number after the original subset in data. This procedure will create a new subset of numbers and after creating this subset, calculate the average. This process must be repeated for all data series [154]. Therefore, when plotting the MA, a line will connect all averages occurring from calculating the moving averages.

A common use of moving average is to smooth out short-term variations and highlight long-term trends or cycles in time series data. The application that uses the moving average will define the edge between short-term and long-term; therefore, the parameters of the moving average will be set accordingly [155]. For example, finance applications often use moving average in technical analysis of financial data such as stock prices returns or trading volumes. Another example of using moving average is in economics to examine a gross national product, employment or other macroeconomic time series. The moving average could be viewed as an example of a low-pass filter that is used in signal processing. However, the moving average has different versions. Therefore in this research, some of them are applied to detect trends in ADLs of an older person or to detect trends in ADWs of a worker in a smart office as presented in this document.

4.3.1.1 Simple Moving Average

Simple Moving Average (SMA) is the unweighted mean of the previous n data. Normally the mean is taken from an equal number of data points on either side of a central value. This ensures that variations in the mean are aligned with the variations in the data rather than being shifted in time. An example of a simple equally weighted running mean for an n-day sample of a recorded data of an activity of an older person who lives in a smart home, is the mean of the previous n days' of a specific activity of this person. If those recorded data are P_t, P_{t-1}, ... $P_{t-(n-1)}$ then Equation 4.2 is used to calculate the SMA.

$$SMA = \frac{P_t + P_{t-1} + ... + P_{t-(n-1)}}{n} \tag{4.2}$$

Each time that new values of SMA are calculated, the old values are dropped out and new values are arrived to make a new SMA, meaning no need to a full summation for this case as it is described in Equation 4.3.

$$SMA_{new} = SMA_{old} + \frac{P_t}{n} - \frac{P_{t-n}}{n} \tag{4.3}$$

The period selected depends on the type of movement of interest, such as short, intermediate, or long-term. The data used in SMA should be centred around the mean. If it is not, then the SMA will lag behind the latest datum point. An SMA could be disproportionately influenced by dropping out an old datum point or by the added new data. When data has periodic fluctuation, then applying SMA for the period will eliminate that variation because the average always contains one complete cycle [155]. Figure 4.3 illustrates the trend in the data set which represents the start time of sleeping using SMA.

4.3.1.2 Cumulative Moving Average

Cumulative Moving Average (CMA) is a method that is used when the user requires the average of all data up until the current datum point; therefore, the data arrives in an ordered datum flow [156]. For example, a carer may want the average of sleeping duration of an older person for particular nights until the current time. For each new calculation of the average of duration at this point,

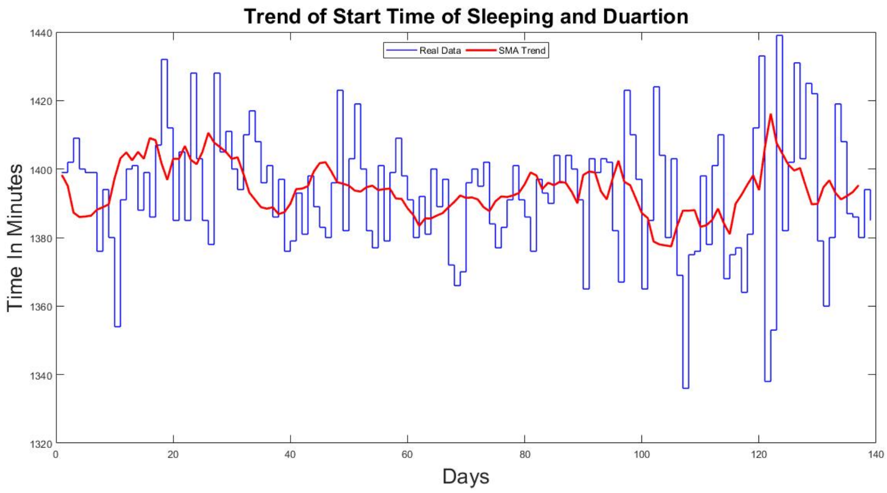

Figure 4.3: Illustrating the trend in the start time of sleeping using SMA.

the CMA can be calculated for all datum points up to this point. Typically CMA uses an equally weighted average of the sequence of n values P_1, P_2,... P_n up to the current point as in Equation 4.4:

$$CMA_n = \frac{P_1 + P_2 + ... + P_n}{n} \tag{4.4}$$

To calculate the average requires all data to be stored in a queue and calculate the sum each time a new datum point arrives, dividing the sum by the number of datum points achieved at the time. However, when a new value P_{n+1} becomes available updating cumulative average can be done simply using Equation 4.5:

$$CMA_{n+1} = \frac{P_{n+1} + n.CMA_n}{n+1} \tag{4.5}$$

Thus, the new datum point cumulative average is equal to the previous cumulative average times n plus the latest datum point and all divided by the number of points received so far $(n + 1)$. At the end when all of the datum points arrive $(n = N)$, the cumulative average will equal the final average. That is;

$$P_1 + P_2 + ... + P_n = n.CMA_n \tag{4.6}$$

and similarly for $n + 1$, it is seen as in Equation 4.7.

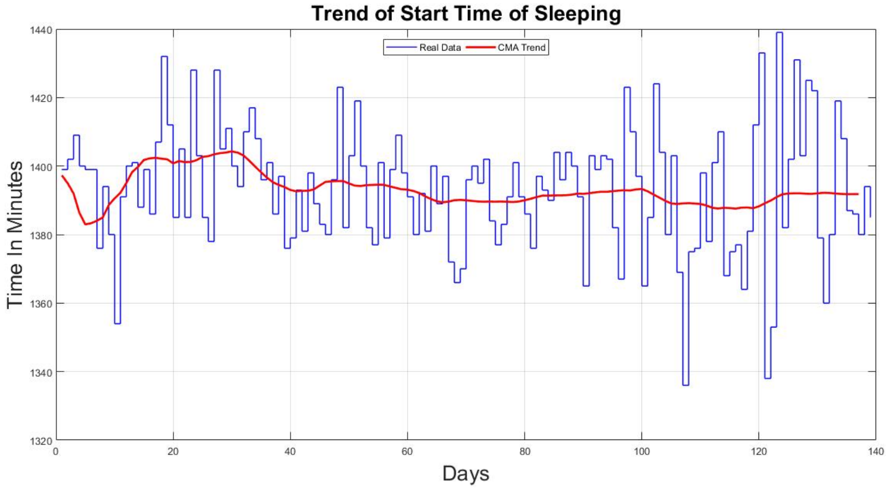

Figure 4.4: Illustrating the trend in the start time of sleeping using CMA.

$$P_{n+1} = (P_1 + ... + P_{n+1}) - (P_1 + ... + P_n) = (n+1).CMA_{n+1} - n.CMA_n \quad (4.7)$$

Finally, to calculate CMA_{n+1}, Equation 4.8 should be used.

$$CMA_{n+1} = \frac{P_{n+1} + n.CMA_n}{n+1} = CMA_n + \frac{P_{n+1} - CMA_n}{n+1} \quad (4.8)$$

Figure 4.4 shows the CMA calculated to represent the trend in the data set that represents the start time of sleeping in a bedroom.

4.3.1.3 Exponentially Weighted Moving Average

Exponentially Weighted Moving Average (EWMA), also known as an Exponential Moving Average (EMA) gives different weights to data points at various positions. The EMA can be an infinite impulse response filter that applies weighting factors which decrease exponentially and never reach zero [153, 157]. The EMA for a series P may be calculated recursively:

$$EMA_1 = P_1 \quad (4.9)$$

for

$$t > 1, EMA_t = \beta.P_t + (1-\beta).EMA_{t-1} \tag{4.10}$$

where:

- β represents the degree of weighting decrease, which is a constant smoothing factor between 0 and 1. A higher β discounts older observations faster.
- P_t is the value at time t.
- EMA_t is the value of the EMA at time t.

EMA_1 could be initialised in diverse ways; the most common way by setting EMA_1 to P_1. It could be initialised by setting EMA_1 to an average of the first 4 or 5 observations. It is very important to initialise EMA_1 because it affects the resultant moving average depending on the β value. Choosing small values of β make the choice of EMA_1 relatively more important than large β values. Higher β values will discount older observations faster. EMA_t as a weighted sum of datum points P_t is presented as [156]:

$$\begin{aligned} EMA_t \;\; = \;\; & \beta.(P_{t-1} + (1-\beta).P_{t-2} + (1-\beta)^2.P_{t-3} + \\ & ... + (1-\beta)^k.P_{t-(k+1)}) + (1-\beta)^{k+1}.EMA_{t-(k-1)}. \end{aligned} \tag{4.11}$$

for any suitable $k = 0, 1, 2, ...$ The weight of the general datum point P_{t-i} is $\beta(1-\beta)^{(i-1)}$. An alternate approach uses P_t in lieu of P_{t1} is described in Equation 4.12

$$EMA_t = \beta.(P_t + (1-\beta).EMA_{t-1}. \tag{4.12}$$

The equation can be expressed in technical terms as in Equation 4.13 to show the steps of EMA towards to the latest datum point, using a proportion of the difference (each time).

$$EMA_{new} = EMA_{old} + \beta.(P_{current} - EMA_{old}) \tag{4.13}$$

where:

- β represents the degree of weighting decrease, which is a constant smoothing factor between 0 and 1. A higher β discounts older observations faster.
- $P_{current}$ is the data value at current time.
- EMA_{old} is the old value of the EMA.
- EMA_{new} is the new value of the EMA.

Expanding out EMA_{old} each time results in a series that shows how the weighting factor on each datum point P_1, P_2, etc. decreases exponentially as shown in Equation 4.14:

$$EMA_{new} = \beta.(P_1 + (1-\beta)P_2 + (1-\beta)^2 P_3 + (1-\beta)^3 P_4 + ...) \tag{4.14}$$

where

- P_1 is current datum point value
- P_2 is previous datum point value
- and so on.

$$EMA_{new} = \frac{\beta \times (P_1 + (1-\beta)P_2 + (1-\beta)^2 P_3 + (1-\beta)^3 P_4 + ...)}{1 + (1-\beta) + (1-\beta)^2 + (1-\beta)^3 + ...)} \tag{4.15}$$

since $1/\beta = 1 + (1-\beta) + (1-\beta)^2 + (1-\beta)^3 + ...$

This is an infinite sum with decreasing terms [158]. Figure 4.5 illustrates the trend when EWMA is applied to the data set representing the start-time of sleeping.

4.3.1.4 Modified Moving Average

Modified Moving Average (MMA) is similar to SMA in the calculation of the first point, however, all subsequent points are calculated by adding the new datum point value and then subtracting the last average from the resulting sum. The

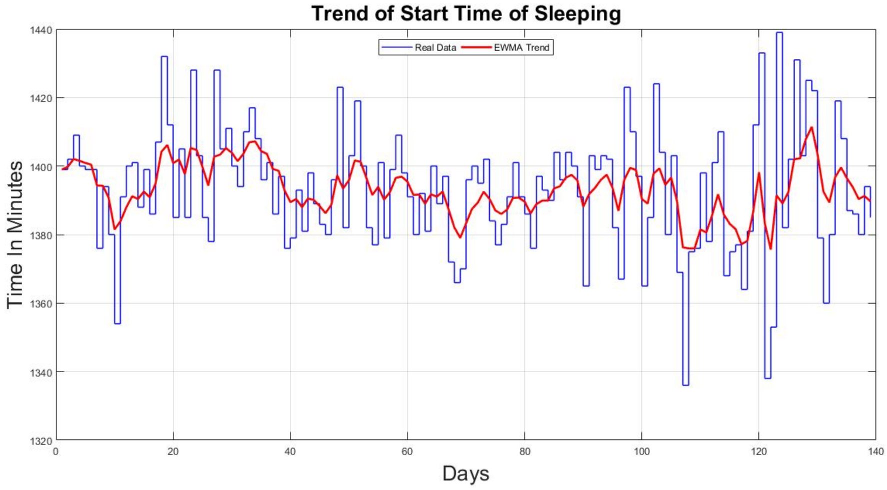

Figure 4.5: Illustrating the trend in the start time of sleeping using EWMA.

difference occurring from last subtraction will be considered as a new point as described in Equation 4.16.

$$MMA_t = MMA_{t-1} + \frac{(P_t - (MMA_{t-1})}{n} \tag{4.16}$$

where MMA_t is the current MMA value, MMA_{t-1} is the previous MMA value and P_t is the current datum point value. In short, this is an EMW with $\beta = 1/n$ [159].

Figure 4.6 illustrates the trend in the data set which represents the start time of sleeping using MMA.

4.3.1.5 Simple Moving Median

When using the moving average to detect trends in a time series, it is susceptible to rare events such as rapid change or other anomalies. Therefore, Simple Moving Median (SMM) over n time points is a more robust estimate of the trend using Equation 4.17

$$SMM = Median(P_t, P_{t-1}, ..., P_{t-n+1}) \tag{4.17}$$

From a statistical point of view, the moving average is an optimal method to detect trends in the time series if the fluctuations about the trend are normally

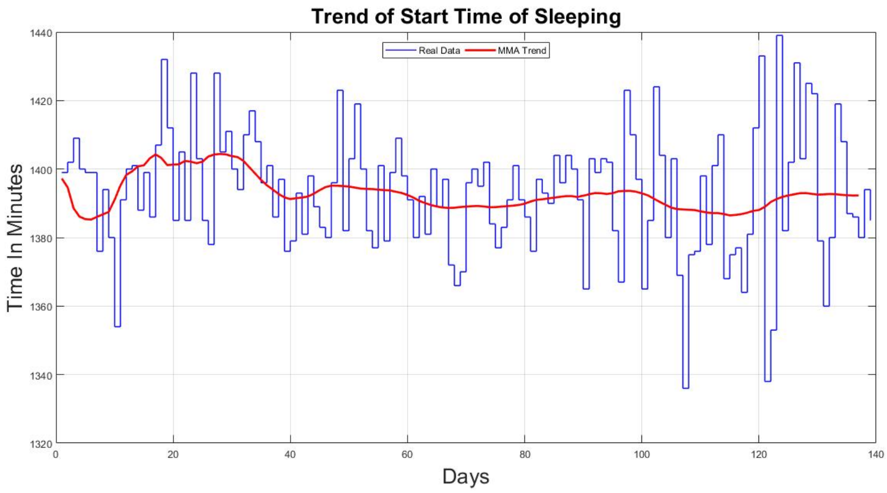

Figure 4.6: Identify the trend in the start time of sleeping using MMA.

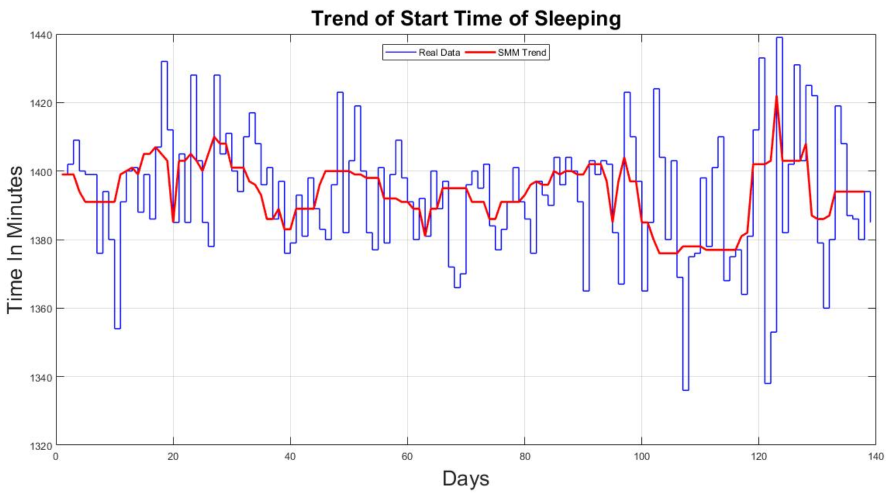

Figure 4.7: Identify the trend in the start time of sleeping using SMM.

distributed. However, if the deviations from trend are very large then the normal distribution does not have a high probability. If the variations are Laplace distributed, then the moving median is statistically optimal [160]. Figure 4.7 illustrates the SMM to detect the trend in the data set which represents the start time of sleeping in a bedroom.

4.3.1.6 n-Point Moving Median

n-Point Moving Median is similar to the SMM and it is calculated based on the number of datum points in each moving window. Therefore, if the original data set has P datum points and the number of data points used in each computation is n, then the number of terms in the moving median sequence will be $P-n+1$ and you can calculate n-point moving median by finding the median of each set of n consecutive points in the series. For example, if you have a sequence of 150 datum point and take their 7 points moving median, the new resulting sequence will have $150-7+1=144$ points.

4.3.2 Seasonal Kendall Test

Seasonal Kendall Test (SKT) is a non-parametric test, which is capable of handling different type of data with different kinds of distribution. The SKT could be used even when there are missing data. The SKT is started by calculating the Mann-Kendall Test (MKT), M. The MKT has three possible results when it is calculated which are: when $M=1$ the change is positive that indicates an upward trend, when $M=-1$ the change is negative which indicates a downward trend, and $M=0$ if data is missing or a tie is present. The Kendall's τ measures the correlation between a chosen parameter and time in Equation 4.18,

$$\tau = \frac{M}{n(n-1)/2} \tag{4.18}$$

where n is the number of samples.

The assumption of the existence of a trend in data is based on calculating τ. Therefore, if $(\tau = 0)$ means no trend otherwise there is a trend [161]. The variance must be calculated using Equation 4.19,

$$\sigma(M_i) = \frac{1}{18}\left(n_i(n_i-1)(2n_i+5) - \sum_{j=1}^{\theta_i} r_{ij}(r_{ij}-1)(2r_{ij}+5)\right) \tag{4.19}$$

where θ_i is the number of tied groups for the i^{th} session and r_{ij} is the number of data points in the j^{th} group for the i^{th} session. For example, if the sequence of measurements of sleeping duration in minutes over 5 months in a specific day is $230, 240, 290, 60, 290, 240, 240, 290, 230$ and we have $\theta_i = 3$ tied groups for that day: $r_{i1} = 2$ for the tied value 230, $r_{i2} = 3$ for the tied value 240, and $r_{i3} = 3$ for the tied value 290.

To analyse ADL's or ADW's data to detect true changes in trends, the seasonal variation must to be removed. The SKT is calculated by the summation of overall M (MKT) for each season as presented in Equation 4.20,

$$K = \sum_{t=1}^{n} M_t \tag{4.20}$$

where n is the number of seasons (samples).

The user can determine the number of seasons (e.g. 24 for hourly sampling, 12 for monthly sampling). The overall variance is calculated by Equation 4.21,

$$\sigma(K) = \sum_{t=1}^{n} \sigma(M_t). \tag{4.21}$$

The results of Equation 4.20 and Equation 4.21 are used to determine Z_s in Equation 4.22,

$$Z_s = \begin{cases} \frac{(K-1)}{(\sigma(K))^{1/2}} & \text{if } K > 0, \\ 0 & \text{if } K = 0, \\ \frac{(K+1)}{(\sigma(K))^{1/2}} & \text{if } K < 0. \end{cases} \tag{4.22}$$

If the absolute value of Z_s is greater than $Z_{1-\alpha/2}$ then the data does not have any specific trend. The trend can be significant if the probability of obtaining K value is less than $0.05(p - value < 0.05)$.

Some issues could arise while using the SKT. For instance, when $K = 0$ the approximation of a large sample is inappropriate and statistically the 100% truth of the null hypothesis (no trend) could be found. To overcome this issue, K is forced to 1 and the $p-value$ is adjusted accordingly. This procedure will maintain the $p - value$, which means no trends over time. Similarly, the estimation of zero slopes are occasionally found to be significant, which is largely driven by the presence of many tied values. Therefore, the tied values should be removed and the estimation of slope should be re-calculated on the remaining values [162].

To increase robustness of the SKT against serial correlation Hirsch and Slack [161] added an estimate of co-variance. Belle and Hughes [163] have developed the test for seasonal homogeneity among the seasons using the Chi-squared test. The SKT results can be false if the individual seasons have heterogeneous trends. For example, if trend in sleeping is increased in one season followed by decreased trend in the next season sequentially then the opposing values of M can cancel each other when K is calculated. To define heterogeneity between individual seasons $H^2_{total}$is broken into two parts, H^2_{homog} and H^2_{trend}, as presented in Equation 4.23

$$H^2_{total} = H^2_{homog} - H^2_{trend} = \sum_{t=1}^{n} Z_t^2 - n\overline{Z}^2 \tag{4.23}$$

where

$$Z_t = \frac{M_t}{\sqrt{\sigma(M_t)}} \tag{4.24}$$

where n is the number of seasons, M_t is the MKT for each season t, and

$$\overline{Z} = \frac{1}{n}\sum_{t=1}^{n} Z_t \tag{4.25}$$

The α critical value is found using a Chi-square distribution table [164]. Therefore, if $X^2_{homog} < \alpha$ critical then there is a common trend to all seasons and the seasons are homogeneous. On the other hand, if $X^2_{homog} > \alpha$ critical then there are different trends for different seasons and seasons are heterogeneous. Therefore, two possible inconsistencies can occur with the SKT when the Chi-Square Test defines heterogeneous seasonal trends. First, if the SKT detects a significant trend through seasons that present different trends then it could be still an overall trend in the data. Second, when the SKT did not find an overall trend, but it could be still a significant trend in individual seasons. To solve this problem, the Mann-Kendall test is used to define trends that may be found in the individual seasons as presented above in this section.

4.4 Moving Average Crossover

The analysis of time series using a moving average crossover occurs when two moving averages are plotted and each of them has a different degree of smoothing. The crossover does not predict the future but it shows the trend's direction. The crossover between two (or more) moving averages is an indicator of trend direction. It is using slower and faster moving averages. The faster-moving average represents a short-term moving average and the slower moving average represents a long or medium term moving average. The slower moving average is moving faster because it considers a small number of datum points in each period and is thus more reactive to each datum point changes and the faster-moving average is estimated slower as it calculates datum points over a long period and is more lethargic. For example, using data of sleeping time duration, for faster-moving average it may be 7, 14 or 28 day period while the slower moving average is medium or long term moving average (e.g. 56, 112 or 224 day period). The moving average as a line in charts is an indicate of trends in data. The crossover will occur when the faster moving average crosses the slower moving average, which is used as a signal of change in trend and can be used to trigger the attention of the carer.

4.5 Prediction Techniques

In this section prediction techniques are presented. These techniques are used in this research to predict the behavioural pattern or the direction of trends in the human behaviour of a user who uses a smart environment based on sensory information. The prediction will help to provide related information of the user's well-being trend. The investigated techniques are:

4.5.1 Activity Prediction Moving Average

The Activity Prediction Moving Average (APMA) has its foundation in the Exponentially Weighted Moving Average (EWMA) to predict the next value of the activity. In contrast to EWMA, it is used to analyse data points by creating a

Algorithm 2 The Prediction Algorithm APMA.

procedure APMA
    Prepare the binary data to be a time series.
    Define training data T.
    Identify the number of data points that need to be predicted m
    Identify the weighted start value of β.
    repeat
        For time t, session m
        repeat
            Calculate EWMA for N data points for all data T
            Calculate the S the start point to predict first ..
            value of N based on historical n data points
            Use S to calculate the next value of N
            if N not accepted **then**
                Modify S and β.
        until all required sessions covered
    until all required time covered

series of averages. More details about EWMA are already provided in Section 4.3.1.3. Based on the EWMA we build a model that can be used to predict the next value of a single activity. The model is tested with our data sets and it shows good results in terms of estimating the next value of the activity. For instance, when it used to predict five days of occupancy duration in an office, the APMA estimates these values of duration. The algorithm 2 illustrates the procedure.

4.5.2 Adaptive-Network based Fuzzy Inference System (ANFIS)

The basic theory of ANFIS model is based on if-then rules and several input-output parameters. ANFIS uses the training, learning algorithms of neural networks [165, 166]. To clarify the explanation, we will assume that the fuzzy inference system has two inputs x and y and one output z and one of the fuzzy models (we will assume a first order Sugeno model). Then the typical fuzzy rule set if-then can be expressed as the following expression:

$$IF\ x\ is\ A1\ and\ y\ is\ B1\ then\ f_1 = p_1x + q_1y + r_1 \tag{4.26}$$

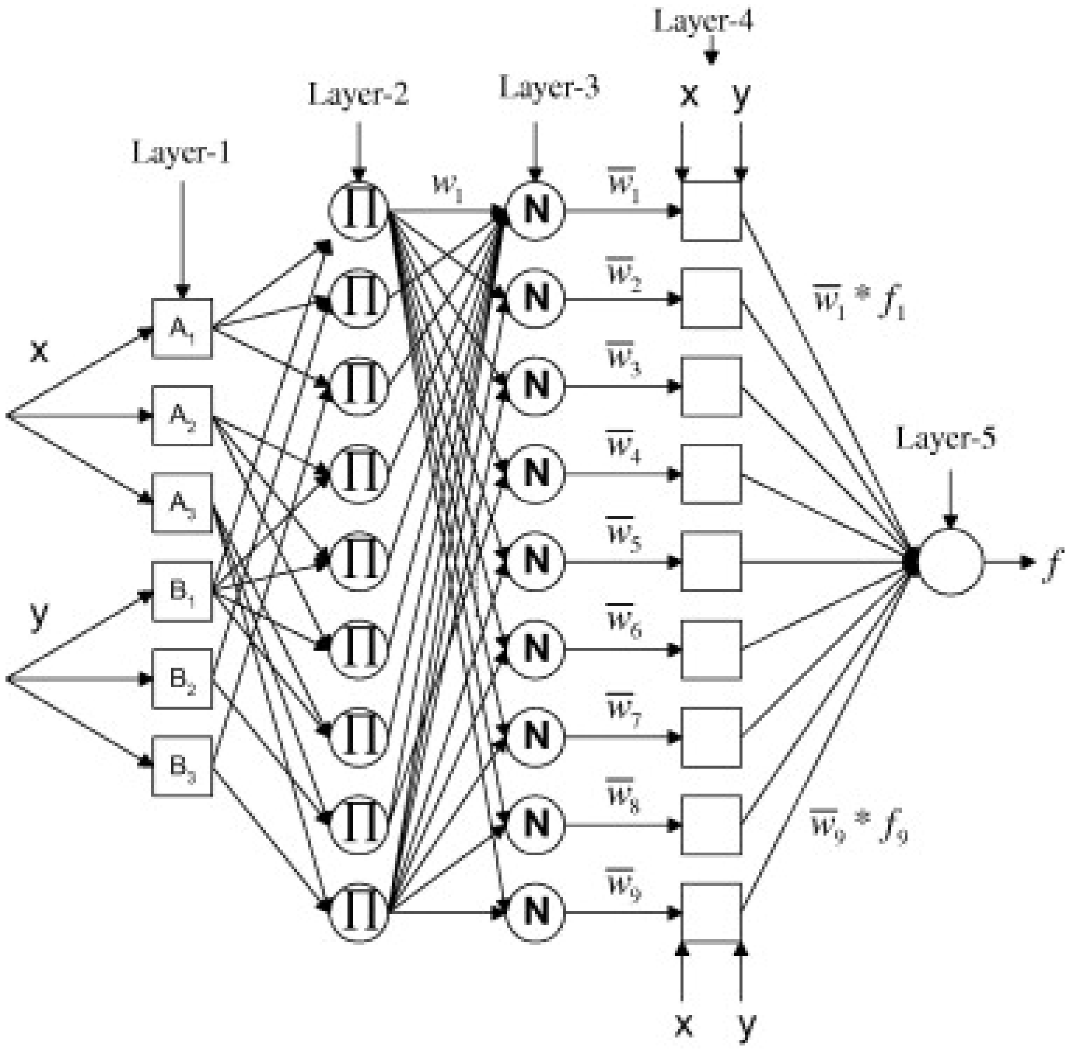

Figure 4.8: ANFIS Architecture Example.

where: p, r, and q are linear output parameters.

Figure 4.8 illustrates the architecture of ANFIS model that has two inputs and one output. The ANFIS model is using five layers and nine ifthen rules as explained below:

- *Layer*$_1$: In this layer Each node i is a square node with a function represented in Equation 4.27

$$O_{1,i} = \mu A_i(x),\ for\ i = 1, 2, 3\ O_{1,i} = \mu B_{i-3}(y),\ for\ i = 4, 5, 6 \tag{4.27}$$

where: x and y are the node i inputs, A_i and B_i are linguistic labels associated with inputs' nodes, O_{1_i} is the membership function of A_i and B_i. Also, it is common to choose $\mu A_i(x)$ and $\mu B_i(y)$ to be a bell-shaped with maximum equal to 1 and minimum equal to 0, which is represented in Equation 4.28

$$\mu A_i(x), \mu B_i(y) = exp(((x_i c_i)/(a_i))^2) \tag{4.28}$$

where: a_i, c_i are premise parameters set.

- $Layer_2$: In this layer each node is shown in a circle labelled with which multiplies the received signals. The output of each node representing the firing strength of a rule. Equation 4.29 illustrate this procedure.

$$O_{2,i} = w_i = \mu A_i(x).\mu B_{i-3}(y)\ for\ i = 1, 2, 3...9 \tag{4.29}$$

- $Layer_3$: The nodes of this layer are shown in circles labelled with N. Each node calculates the ratio of the ith rules. Equation 4.30 explains the layer's processing.

$$O_{3,i} = \overline{w}_i = w_i/(w_1 + w_2 + ... + w_9)\ for\ i = 1, 2, 3...9 \tag{4.30}$$

- $Layer_4$: Equation 4.31 shows the function of each node in this layer. Also, every node is a square node.

$$O_{4,i} = \overline{w}_i.f_i = w_i.(p_1 x + q_1 y + r_1)\ for\ i = 1, 2, 3...9 \tag{4.31}$$

where: w_i is the output and p_i, q_i, r_i are consequent parameters.

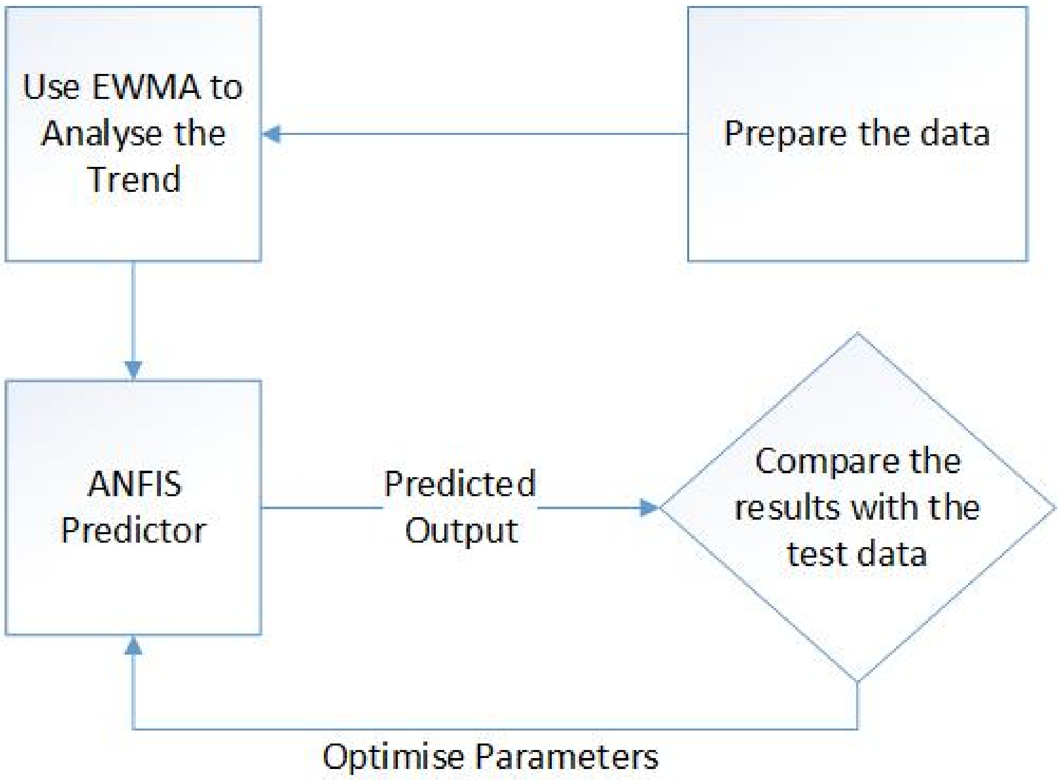

Figure 4.9: The schematic diagram of the proposed model to predict trends using ANFIS.

- $Layer_5$: The last layer is represented by circle and labelled by $\sum$. It is computing the final output as shown in Equation 4.32

$$O_{5,i} = overall\ output\ = \sum_i \overline{w}_i.f_i = \frac{\sum_i w_i.f_i}{\sum_i w_i} \tag{4.32}$$

The proposed model based on ANFIS is used to predict trends in the data sets. Figure 4.9 illustrates the schematic diagram of the proposed model. The model starts with preparing our binary data sets to be time series, then the data is smoothed using EWMA technique. The output data from using EWMA will be passed to the ANFIS model to predict the data that will represent the human activities. The predicted values will be examined against the training data sets to adapt the ANFIS rules to get better results.

4.5.3 Moving Average Convergence/Divergence

Moving Average Convergence/Divergence (MACD) is an indicator used in technical analysis of time series data. It was created by Gerald Appel in the late 1970s [167]. It is proposed to detect trends and their changes in the strength, direction, momentum, and duration. The MACD indicator is based on three time series calculated from historical data. These three series are the MACD series, the "signal" (average) series, and the "divergence" series which is the difference between the signal and MACD. The MACD series is the difference between a "fast" (short period) exponentially weighted moving average (EWMA), and a "slow" (longer period) EWMA of the data series. The average series is an EWMA of the MACD series itself [168].

The MACD indicator depends on three parameters, namely the time measurements of the three EWMAs. The MACD indicator calculated using the notation "MACD(a,b,c)" where the MACD series is the difference between two EWMAs with representative times a and b, and the average series is an EWMA of the MACD series with representative time c. These parameters are usually measured in datum points. For example, to calculate MACD using duration of sleeping time, values such 14, 28 and 7 days could be used, $MACD(14, 28, 7)$.

The MACD and the signal are usually presented as continuous lines in the chart, while the divergence is displayed as a bar graph (histogram)[169]. The fast EWMA responds more quickly than the slow EWMA to recent changes in data. Therefore, the MACD series can indicate changes in the trend of data by comparing EWMAs of different periods. The divergence series can reveal subtle shifts in the data's trend [170].

The MACD is used to measure changes in the trend of studied data. It is using the difference between the series and its average to detect subtle shifts in the strength and direction of trends. It needs to pay special attention when the MACD line crosses the signal line or when the MACD line crosses the zero axis. Importance is also found in the disagreement between the MACD line or the difference line and the real data (specifically, higher highs or lower lows on the data series that are not matched in the indicator series).

The "signal-line crossover" occurs when the MACD line and the signal lines

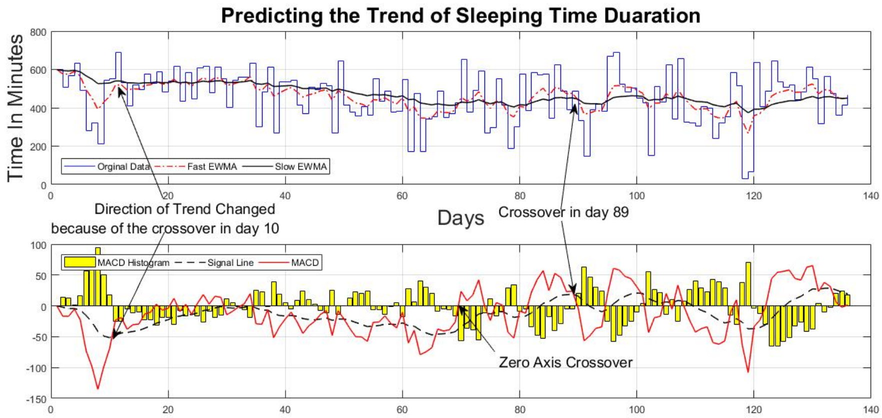

Figure 4.10: Predicting the direction of trend in sleeping time duration.

cross each other; that happens when the divergence (the bar graph) changes sign. The interpretation of this crossover is that when the MACD line moves upward or downward it is taken as an indication that the trend in the studied data is about to accelerate in the direction of the crossover.

The "zero crossover" occurs when the sign of MACD is changed and the MACD line crosses the horizontal zero axis. The crossing occurs when there is no difference between the fast and slow EWMAs. The crossover provides evidence of the trend direction change but less confirmation of the signal line crossover.

Divergence has two types: positive and negative divergence. The "positive divergence" or "bullish divergence" occurs when the studied data records a new low value and the MACD does not confirm with its own new low value. The "negative divergence" or "bearish divergence" occurs when the studied data records a new high value and the MACD does not confirm with its own new high value. The divergence could occur on the MACD line and/or the MACD histogram [171, 168]. Figure 4.10 illustrates a sample of using MACD to predict the duration of trends in a human behaviour data set representing the duration of sleeping time for an older person.

4.5.4 Relative Strength Index

The Relative Strength Index (RSI) is an indicator used in the analysis of financial data. It was proposed to chart the strength or weakness of a stock based on the closing prices of a recent trading period compared to the historical prices. The RSI is a momentum indicator uses to measure the velocity and magnitude of directional datum points movements. The RSI calculates the ratio of higher values to lower values in studied data. The momentum is the rate of the rise or fall in datum points. The datum point that has more or stronger positive changes has a higher RSI than the datum point which has more or stronger negative changes [172]. RSI starts by calculating an upward change U or downward change D of each data period as presented in equations 4.33, 4.34. U occurs when the current datum point being higher than the previous datum point and D occurs when the current datum point being lower than the previous datum point. However, if the current datum point is the same as the previous, then both U and D are zero.

$$\begin{aligned} &D = datum_{previous} - datum_{current} \\ &U = 0 \end{aligned} \tag{4.33}$$

$$\begin{aligned} &U = datum_{current} - datum_{previous} \\ &D = 0 \end{aligned} \tag{4.34}$$

The average U and D are calculated using one of the moving average techniques mentioned earlier. The ratio of these averages is the relative strength or relative strength factor which is described in equation 4.35.

$$RS = \frac{MA(U,n)}{MA(D,n)} \tag{4.35}$$

If the average of D values is zero, then relative strength factor is converted to a relative strength index between 0 and 100 as shown in Equation 4.36.

$$RSI = 100 - \frac{100}{1+RS} \tag{4.36}$$

The moving averages should be properly initialised using the first n values in the data series. The plotted RSI is presented on a graph above or below the data chart. The typical upper line of the indicator is plotted at 70, a lower line at 30, and a dashed mid-line at 50. The level of the RSI is a measure of the studied data's recent change strength. The RSI's slope is directly comparative to the velocity of a change in the trend and the RSIs distance travelled is proportional to the magnitude of the move.

Traditionally, when RSI readings become more than the 70 level, they are considered to be in over-level territory, and when the readings became lower than the 30 level, they are considered to be in lower-level territory. If the readings are in-between the 30 and 70 level then they are considered as neutral and no trend when the sign at 50 level.

The divergence between RSI and data is a very strong indication of an imminent turning point. It is called "bearish divergence" when a new high datum occurs and the RSI makes a lower high, thus failing to confirm. "Bullish divergence" occurs when data makes a new low but RSI makes a higher low [173, 174].

4.6 Discussion

This chapter gives an overview of different techniques used in this research as solutions to identify trends in time series and to predict new data values which represent ADL or ADW. SMA, normally displayed as a smoothed line, is less prone to respond to slight and temporary changes of data. It is good for long time frames, such as daily or weekly charts. It is slower to respond to rapid data changes. It ignores complex relationships in data. CMA, which displays as a smoothed line, is a simple method to get an average of all the data up until the current datum point. It is slower to respond to rapid data changes. MMA responds more quickly to data changes. It is more vulnerable to false signals. SMM is sensitive to rapid changes and anomalies in data. EWMA responds more quickly to data changes and it is helpful to deal with data within a shorter time frame. It is more vulnerable to false signals. Finally, SKT can deals with seasonality in data. It will give satisfactory results even if there is a missing data. It needs to deal with each season first, which can lead to false results if the individual seasons have heterogeneous trends.

Some predictive techniques used in this research were also discussed. APMA is based on EWMA and it can be used to predict data or the direction of the trend. ANFIS is widely used in different disciplines. In this research it is used after using EWMA for the data, then it is used for prediction. MACD is used to measure the momentum and predict the trend's direction. It has the ability to foreshadow moves in the underlying data. It does not have boundaries to bind or limit its movement. MACD represents the difference between two moving averages, therefore it could still have some lags. RSI can be used to identify outliers in data by using the indicators over-level and lower-level. It is used to predict the direction of trends. However false signals still could occur. Chapter 5 will describe the implementation of these techniques and compare between them. The techniques used are evaluated using data sets collected from ADLs or ADWs of smart environments users.

Chapter 5

Trend Analysis and Prediction Techniques for One Activity

5.1 Introduction

To identify the progressive changes in human behaviour activities, it is essential to identify initially the changes for each single activity. This could be even more simplified and the problem in hand is defined as identifying the progressive changes for a single sensor at a time. For example, to measure the progressive changes in sleeping time could be represented as progressive changes measured from a PIR in a bedroom or a bed pressure sensor. In this chapter existing trend analysis techniques mentioned in Chapter 3 are applied to data sets explained in Chapter 4. Later, in this chapter, Human Behaviour Momentum Indicator is introduced which is a more suitable measure to identify the progressive changes for one activity.

5.2 Implementing Existing Trend Analysis Techniques

The problem of detecting trends in the pattern of human behaviour has increasingly attracted attention. Using trend analysis techniques will help in detecting abnormal behaviour which is the identification of previously unknown patterns.

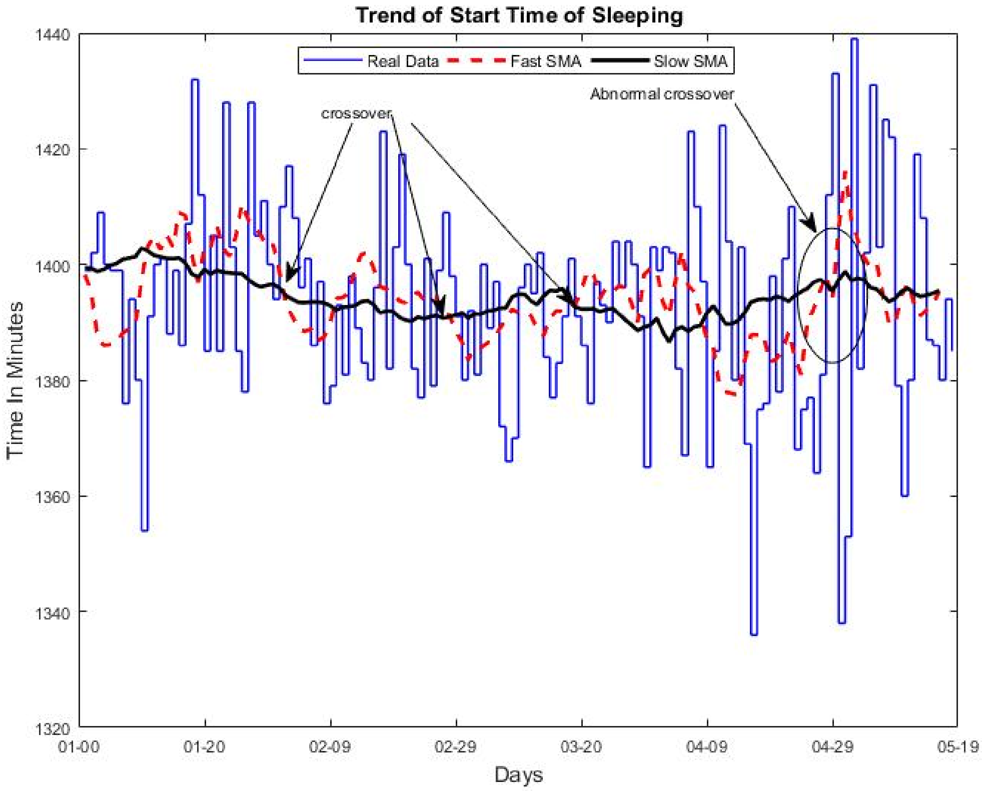

Figure 5.1: Trend of sleeping start time using 7 and 28 data points to compute SMA.

The problem is particularly difficult because what constitutes an abnormal behaviour can greatly differ depending on the task at hand. In the following sections, implementing trend analysis techniques on human behaviour data sets described in Chapter 4 are investigated.

5.2.1 Simple Moving Average

The calculation of Simple Moving Average is conducted for extracted data from monitoring an elderly person who lives in a smart home. The proposed algorithm that is used to implement the SMA shows good results of detecting trends. To apply the SMA, three pairs of numbers representing fast and slow SMA are chosen. Each number states the starting time of sleeping in days. First two pairs

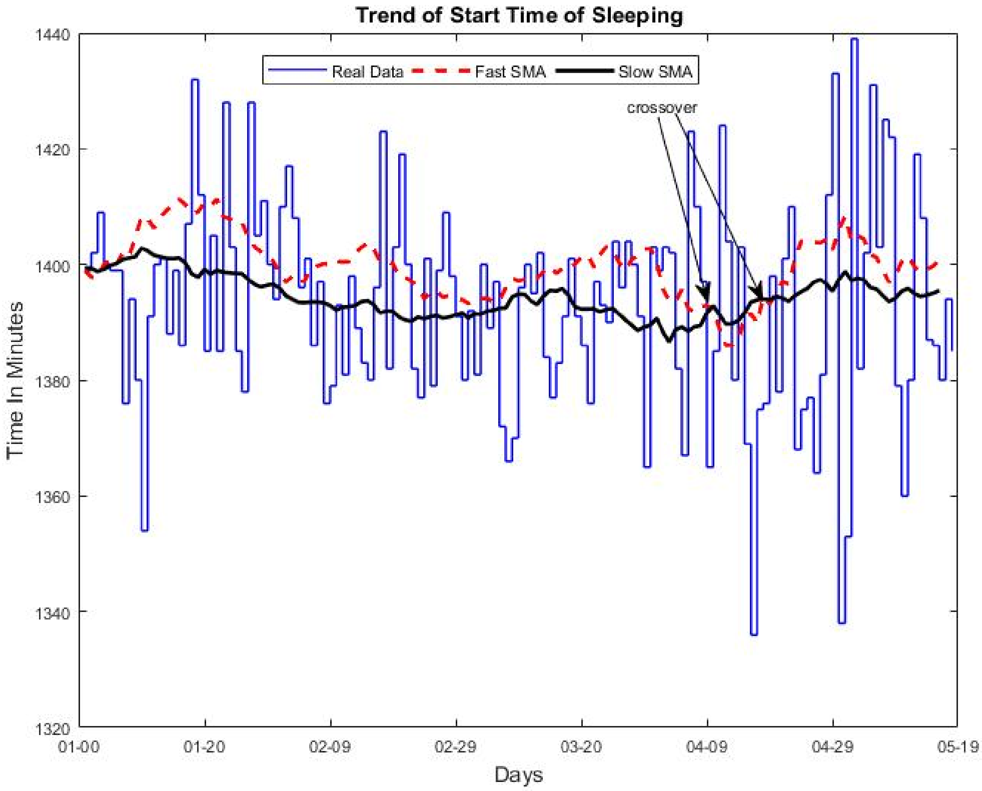

Figure 5.2: Trend of sleeping start time using 14 and 28 data points to compute SMA.

are (7, 14) datum points. The 7 days values are used to compute the fast SMA and the slow SMA is computed using 14 datum points' values. Second pairs are 14 and 28 for fast and slow SMA. The last pairs are 14 for fast SMA and 56 for Slow SMA.

When a new datum comes available the old data is dropped, which will cause the average to move along the time scale. For example using 7 datum points values, the first value of the SMA is calculated by covering the first seven data values. The second SMA's value is calculated by dropping the first data point and adds the new data point (in this case number eight in original data). The third SMA's value continues by dropping the first data point from previous data that is used to calculate the second SMA value (second data point in real data) and adding the new data point (in this case the nine number in original data).

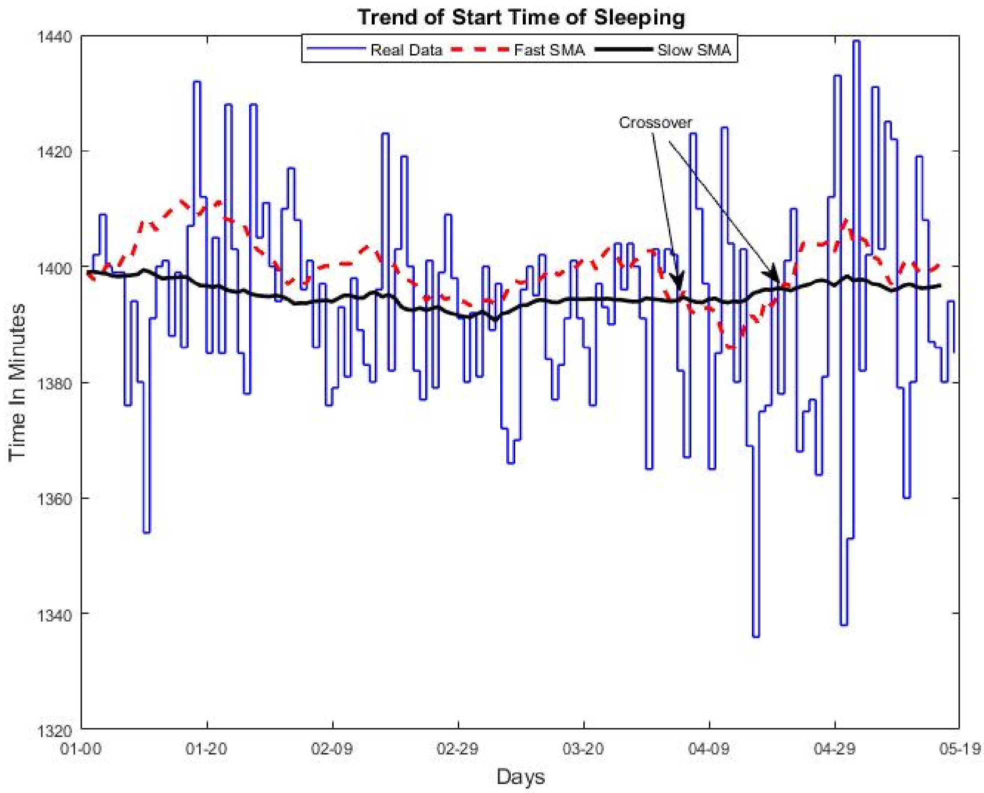

Figure 5.3: Trend of sleeping start time using 14 and 56 data points to compute SMA.

It is noticed that using the first pair (7, 14) is more sensitive to changes happen in original data, and crossover occurs each time change of data up or down happen as it shown in Figure 5.1. However, the second pair (14, 28) has less crossover but it is clearer to show the trend as it is presented in Figure 5.2. The third pair (14, 56) which is shown in Figure 5.3 gives similar results to the second pair with more smooth line for the slow SMA. Moreover, the SMA is shown that it is possible to detect abnormal crossover which represents abnormal activities.

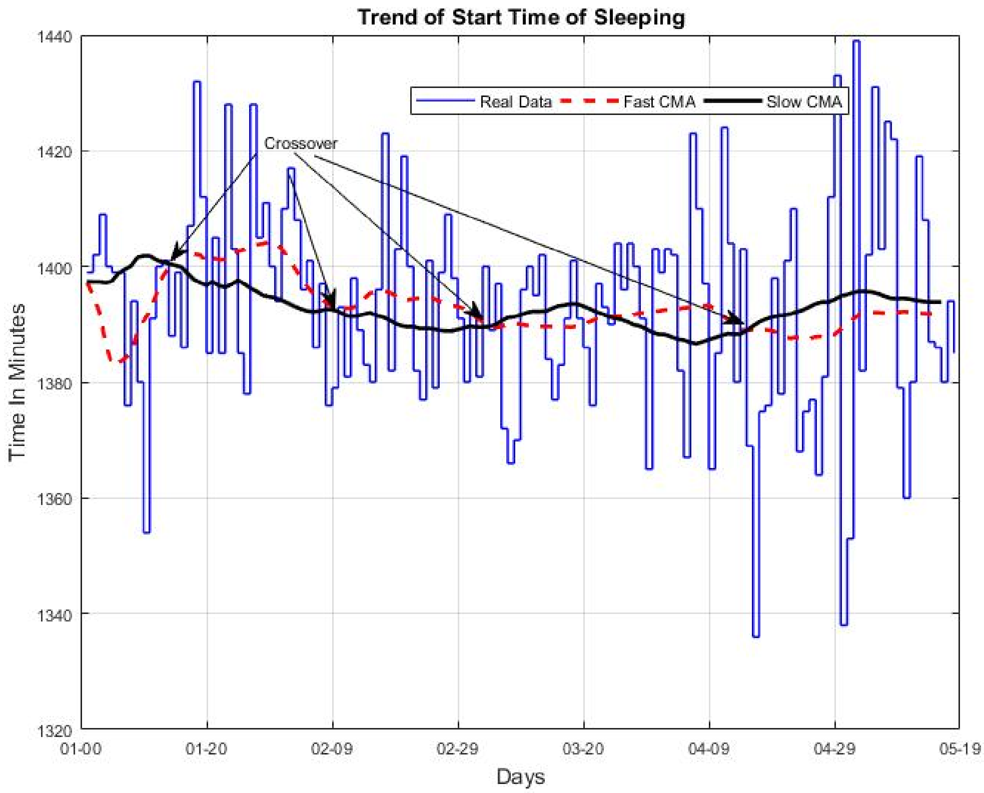

Figure 5.4: Trend of sleeping start time using 7 and 28 data points to compute CMA.

5.2.2 Cumulative Moving Average

The Cumulative Moving Average is calculated for the same date mentioned in the earlier section. The CMA algorithm shows its ability to detect trends in ADLs' data. However, to apply the CMA three pairs of numbers to represent the fast and slow CMA are chosen. Each number states the starting time of sleeping in days. The procedure for calculating CMA using the fast and slow movements is similar to what is presented in Section 5.2.1 to calculate SMA. The data that are used in this example extracted from the original data to represent the start time of sleeping in the bedroom.

Using the first pair (7, 14) is shown that the CMA is more sensitive to changes happen in original data in when it is in fast version, and crossover occurs each

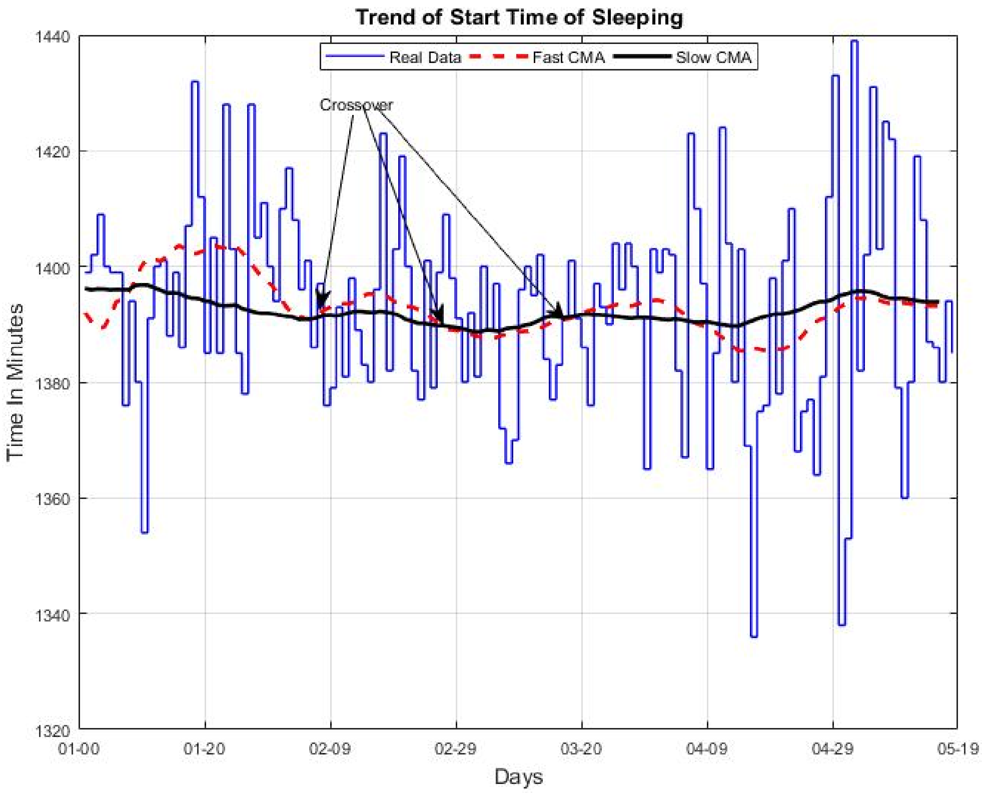

Figure 5.5: Trend of sleeping start time using 14 and 56 data points to compute CMA.

time big change of data up or down happens as it shown in Figure 5.4. However, the second pair $(14, 28)$ has nearly the same crossovers but it is clearer to show the trend. The third pair $(14, 56)$ which is shown in Figure 5.5 gives similar results to the second pair with more smooth line for the slow CMA.

5.2.3 Modified Moving Average Algorithm

The used data in this example to test the modified moving average are extracted from the original data to represent the start time of sleeping in the bedroom. The proposed MMA algorithm can detect trends in the studied data. The implementation of the MMA is based on choosing three pairs of numbers to represent the fast and slow MMA. Each number states the starting time of sleeping in days.

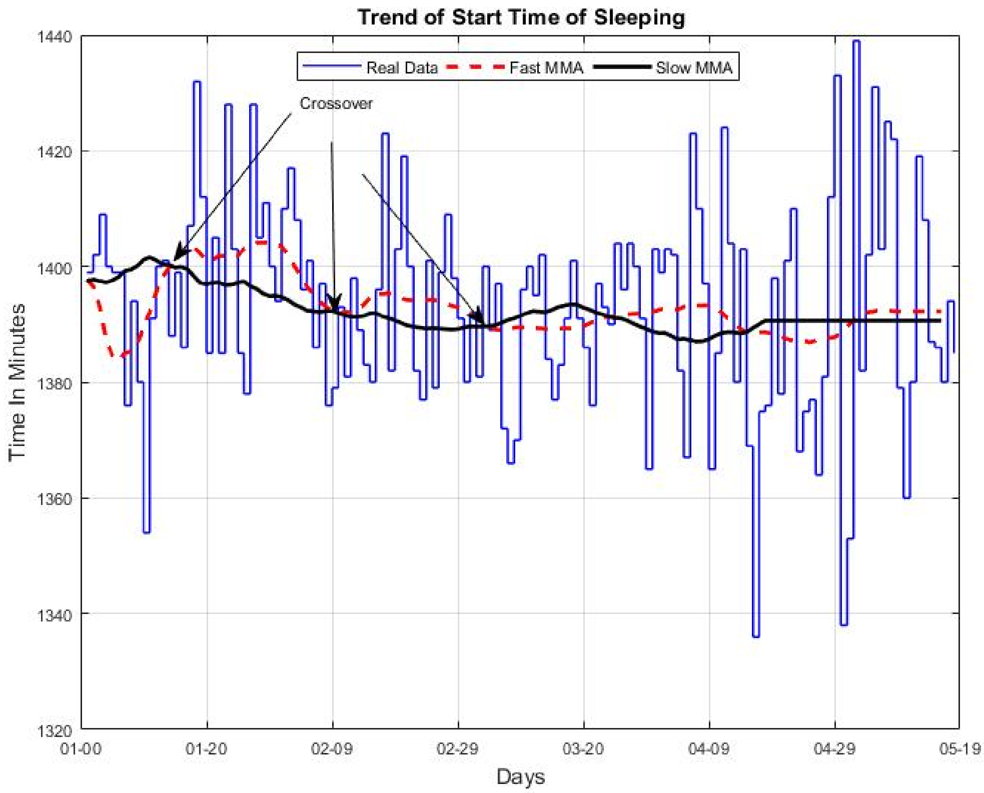

Figure 5.6: Trend of sleeping start time using 7 and 28 data points to compute MMA.

The procedure for calculating MMA using the fast and slow movements is similar to what is presented in Section 5.2.1 to calculate SMA.

In this test the using MMA with first pair (7, 14) is shown that the fast MMA is more sensitive to changes that may happen in original data, and crossover occurs when data changes gradually up or down happen as it shown in Figure 5.6. However, the second pair $(14, 28)$ has nearly the same crossovers but it is clearer to show the trend. The third pair $(14, 56)$ which is shown in Figure 5.7 gives similar results to the second pair with more smooth line for the slow MMA.

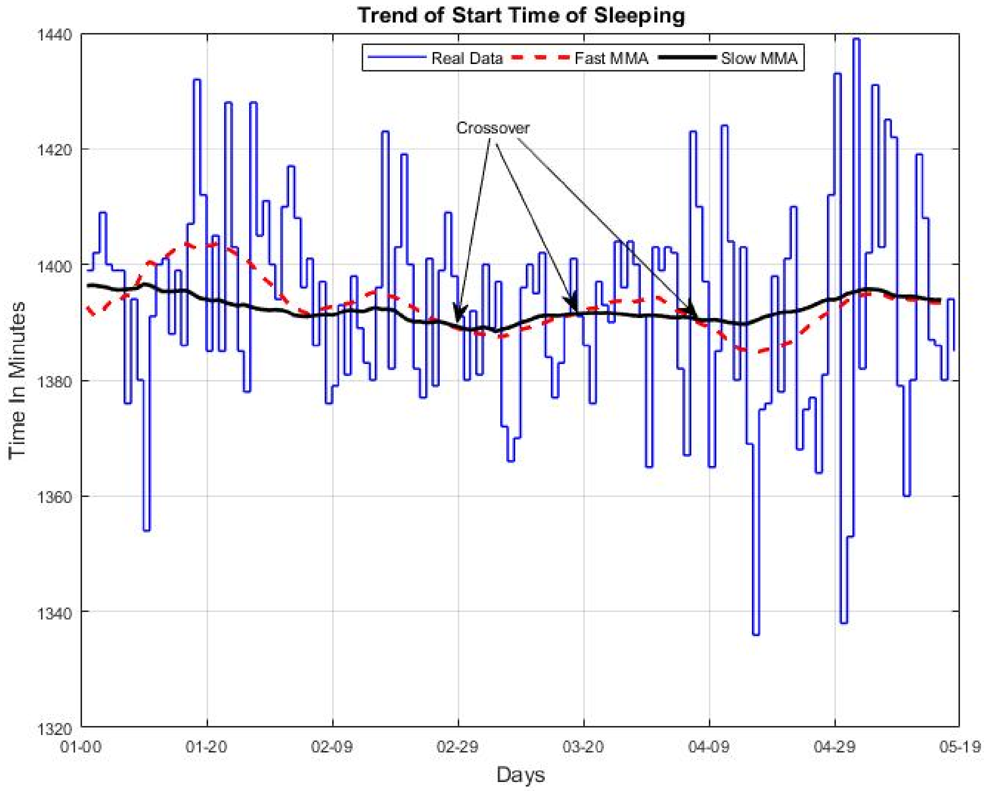

Figure 5.7: Trend of sleeping start time using 14 and 56 data points to compute MMA.

5.2.4 Exponentially Weighted Moving Average

The proposed Exponentially Weighted Moving Average is tested using data that are extracted from the original data to represent the start time of sleeping in the bedroom. The implementation of the EWMA is done by choosing three pairs of numbers to represent the fast and slow EWMA. Each number performs the starting time of sleeping in days. The procedure for calculating EWMA using the fast and slow movements is similar to what is presented in Section 5.2.1 to calculate SMA.

The first pair of numbers are (7, 14), which are used to test the EWMA, they show that the fast EWMA is more sensitive to changes that may happen in original data, and crossover occurs when data changes happens gradually up

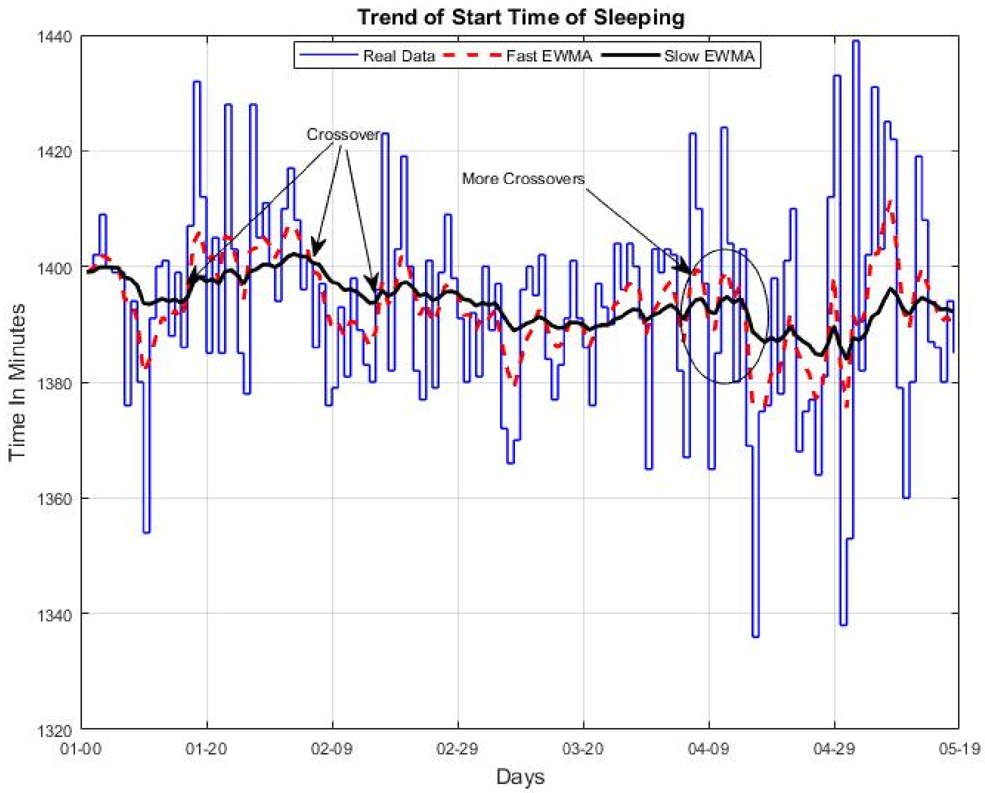

Figure 5.8: Trend of sleeping start time using 7 and 28 data points to compute EWMA.

or down as it is shown in Figure 5.8. However, the second pair $(14, 28)$ has less crossovers but it is clearer to show the trend as it is presented in Figure 5.9. The third pair $(14, 56)$ gives similar results to the second pair with more smooth line for the slow EWMA. The EWMA technique is shown the abnormal crossovers which may represent abnormal activities.

5.2.5 Seasonal Kendall Test

Generally, focusing on finding out a long-term patterns and trends in elderly people activities are key areas of concern; they are necessary to estimate the progress or deterioration in these activities which could help the medical assistant or caregiver. The Seasonal Kendall Test (SKT) is used to find trends in data

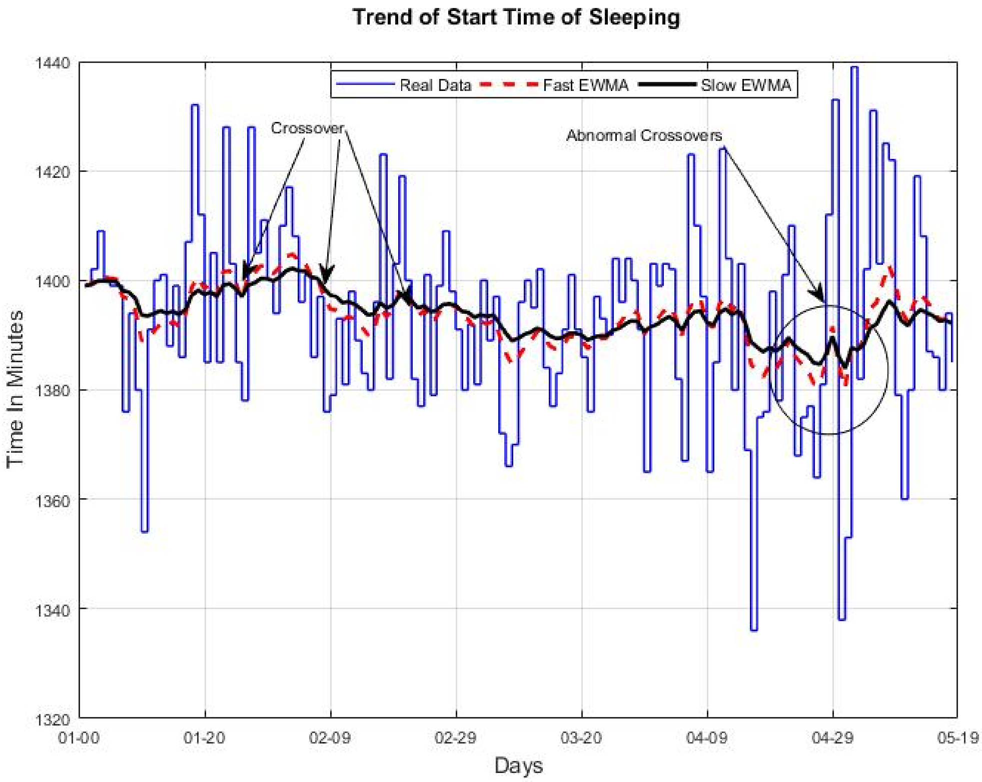

Figure 5.9: Trend of sleeping start time using 14 and 28 data points to compute EWMA.

collected from ADL. The SKT detects the found trends in all data sets that are used in this research. For example, there is a decreasing trend in the sleeping pattern and an increasing trend in visiting the bathroom during sleeping time. Figure 5.10-a for real data shows an example of the decreasing trend of sleeping pattern and the duration time in sleeping in bedroom, Figure 5.10-b shows a decreasing trend of sleeping pattern in simulated data.

Tables 5.1 and 5.2 give the values of Chi Trend, which are exceeded with probability α. That means there is a common trend for all seasons with critical values of each Chi Trend based on value of α. However, Table 5.1 is summarising the value of Chi trends founded in the real and simulated data for the sleeping pattern, as well it gives a comparison between the values of both data to the index values. Also, Table 5.2 is summarising the value of Chi trends founded in

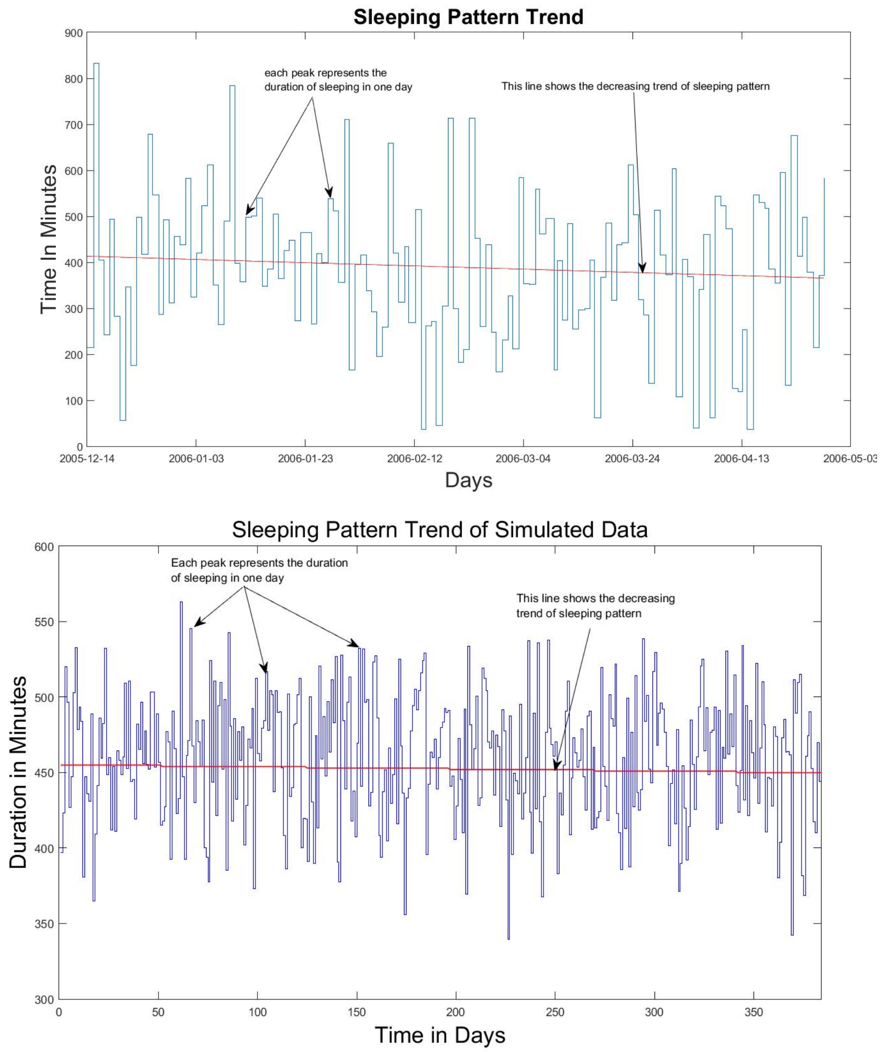

Figure 5.10: Trend of sleeping pattern in a) real data b) simulated data.

the real and simulated data for using the bathroom in sleeping time (distracting the sleeping) pattern, as well it gives a comparison between the values of both data to the index values.

5.3 Implementing Prediction Techniques

In this section the predictive techniques as solutions to the binary time series prediction representing the ADLs or ADWs are presented. Different prediction techniques are used to predict the trends or the direction of trends in time series data representing the human behaviour of the inhabitants in their smart environment. Several experiments are devised, executed, and evaluated using different prediction techniques. In the following section the explanation of these techniques is presented.

5.3.1 Activity Prediction Moving Average

Activity Prediction Moving Average (APMA) is proposed as a predictor model. It is tested using different type of data sets that are extracted from the original data to represent the duration of sleeping in the bedroom and the duration of occupying a smart office. The EWMA is used to detect the trend in the data first, then it is used to predict extra data based on the data that are occurred from the trend. Two pairs of numbers are used to represent the fast and slow EWMA. This method will allow to see the trends of data in different degree of smoothing. In prediction process, it is very important to adjust the weighted value to gain the best results. The other fact in prediction is the smallest number of prediction days the better results will gain.

The first experiment was done with the data of sleeping duration. The (7, 28) data points are used to compute EWMA. Then the APMA is used to predict the trend of last 7 days. Comparing between the results of using fast and slow EWMA, they show that the fast EWMA is more sensitive to changes that may happen in original data, and crossover occurs when data changes happens gradually up or down as it is shown in Figure 5.11. However, the same procedure is

Table 5.1: Values of Chi Trend, which are exceeded with probability α.

α	0.5	0.1	0.05	0.025	0.01	0.001
INDEX	0.45	2.71	3.84	5.02	6.64	10.8
Simulated data	0.45	2.71	3.84	5.02	6.64	10.8
Real data		2.71	3.84	5.02	6.64	10.8
Chi Trend in Simulated data	0.403	0.403	0.403	0.403	0.403	0.403
Chi Trend in Real data	–	0.778	0.778	0.778	0.778	0.778

used with the office's data. The (5, 20) data points are used to compute EWMA and then use the APMA to predict 5 days. With the same concept that is used in the first experiment Figure 5.12 shows the results.

On the other hand, by comparing the EWMA data and the APMA data it can be seen significant results that obtained by using APMA. The results in Tables 5.3, 5.4 show how close the predicted data to the original one.

Table 5.2: Values of Chi Trend, which are exceeded with probability α.

α	0.5	0.1	0.05	0.025	0.01	0.001
INDEX	0.45	2.71	3.84	5.02	6.64	10.8
Simulated data	–	2.71	3.84	5.02	6.64	10.8
Real data		–	–	5.02	6.64	10.8
Chi Trend in Simulated data	–	1.79	1.79	1.79	1.79	1.79
Chi Trend in Real data	–	–	–	5.01	5.01	5.01

Table 5.3: A sample of results of using EWMA and APMA with data of the Bedroom.

EWMA data represent total duration in minutes	**APMA data represent total duration in minutes**
514.70	511.72
524.36	523.21
472.52	477.13
495.05	493.42
489.98	490.29
457.76	460.72
446.67	447.95

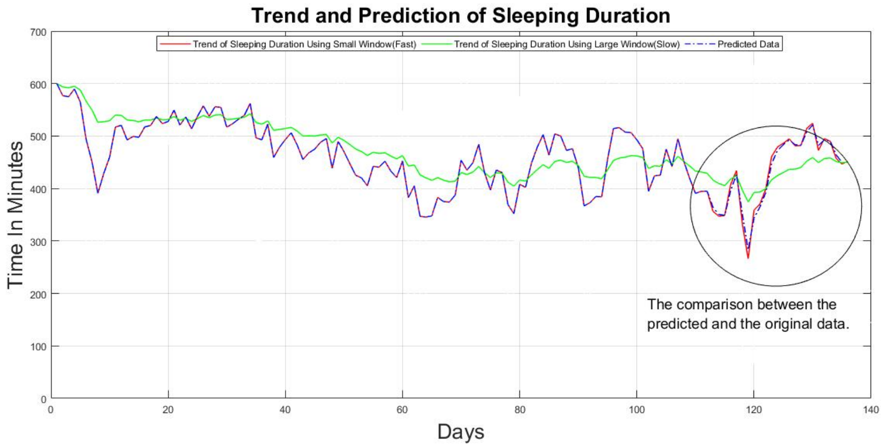

Figure 5.11: Trend and prediction of sleeping duration using EWMA.

5.3.2 Adaptive-Network based Fuzzy Inference System (ANFIS)

In this work, ANFIS is used to predict a value of activity based on historical data set. As it shown in the proposed model, it is essential to prepare the data to be used as a time series. The second step is to use trend analysis technique to smooth each data set and get its trend; then ANFIS is used to predict the new values of data points based on the smoothed data sets. The variants of the algorithm used in the study are different numbers of membership functions and different types of function (i.e "gbellmf", "gauss2mf"). Different data sets are

Table 5.4: A sample of results of using EWMA and APMA with the Smart Office's data.

EWMA data represent total duration in minutes	**APMA data represent total duration in minutes**
384.80	384.36
396.39	393.99
406.01	403.61
390.24	392.91
396.67	395.92

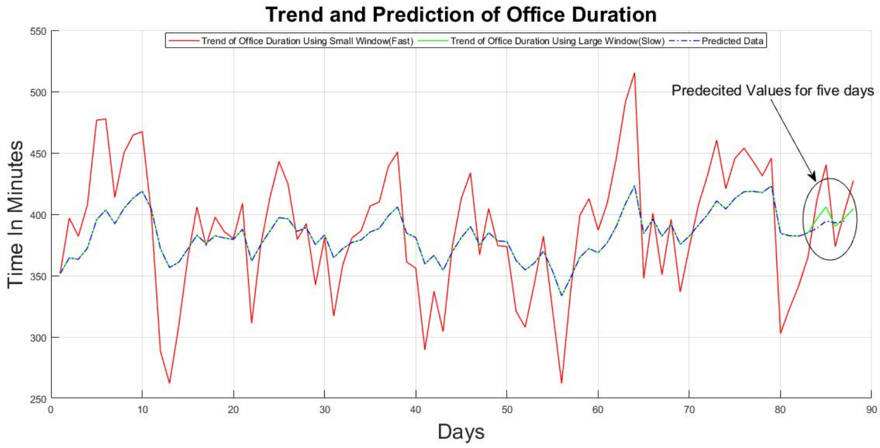

Figure 5.12: Trend and prediction of an office occupying using EWMA.

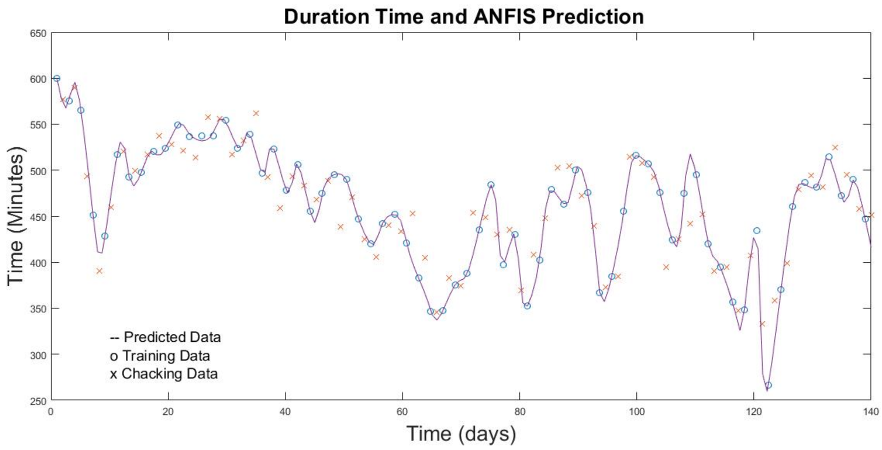

Figure 5.13: Predicted duration time using ANFIS.

used in this work too.

Figure 5.13 represents a sample of results using ANFIS to predict the data using 30 bell shape membership functions. The data set used here is for the duration of sleeping time in a bedroom. Table 5.5 presents a sample of results, this table has two columns; one for actual data and the other is for predicted data. In addition, the root-mean squared (RMS) is used to compare predicted and actual values for model validation.

5.3.3 Human Behaviour Momentum Indicator (HBMI)

To identify changes and ultimately to forecast changes in the ADL or ADW, a new indicator, Human Behaviour Momentum Indicator (HBMI), is introduced. The HBMI is proposed to identify the changes of the activity's behaviour. The model has three phases. First phase is designed to prepare the binary data and convert it to a time series data. The second phase has two tasks, the first task is to examine the converted data and handle it to be standardised. The second task is applying MACD to the available data sets. The third phase is to calculate the indicator, which has two tasks too, the first task is to calculate the index of data and second task is to calculate the boundaries. In other words, the index will be centralised to one hundred (100) that is because the indicator will present the percentage of the changes in the activity in each data point and it will have three boundaries on both sides of the centre value (100). These boundaries will represent three main categories; normal, worrying and abnormal. The following steps show the process of calculating the indicator and the boundaries.

- Calculate

$$Index_i = \frac{Act_i}{Act_{i-n}}100. \tag{5.1}$$

  where: i=1,2,3,.. and n the number of data points that are used to calculate EWMA

Table 5.5: A sample of results of using ANFIS with dataset of the Bedroom.

EWMA data represent total duration in minutes	**ANFIS data represent total duration in minutes**
451.65	479.12
494.67	492.18
481.95	489.84
457.76	458.02
443.10	451.65

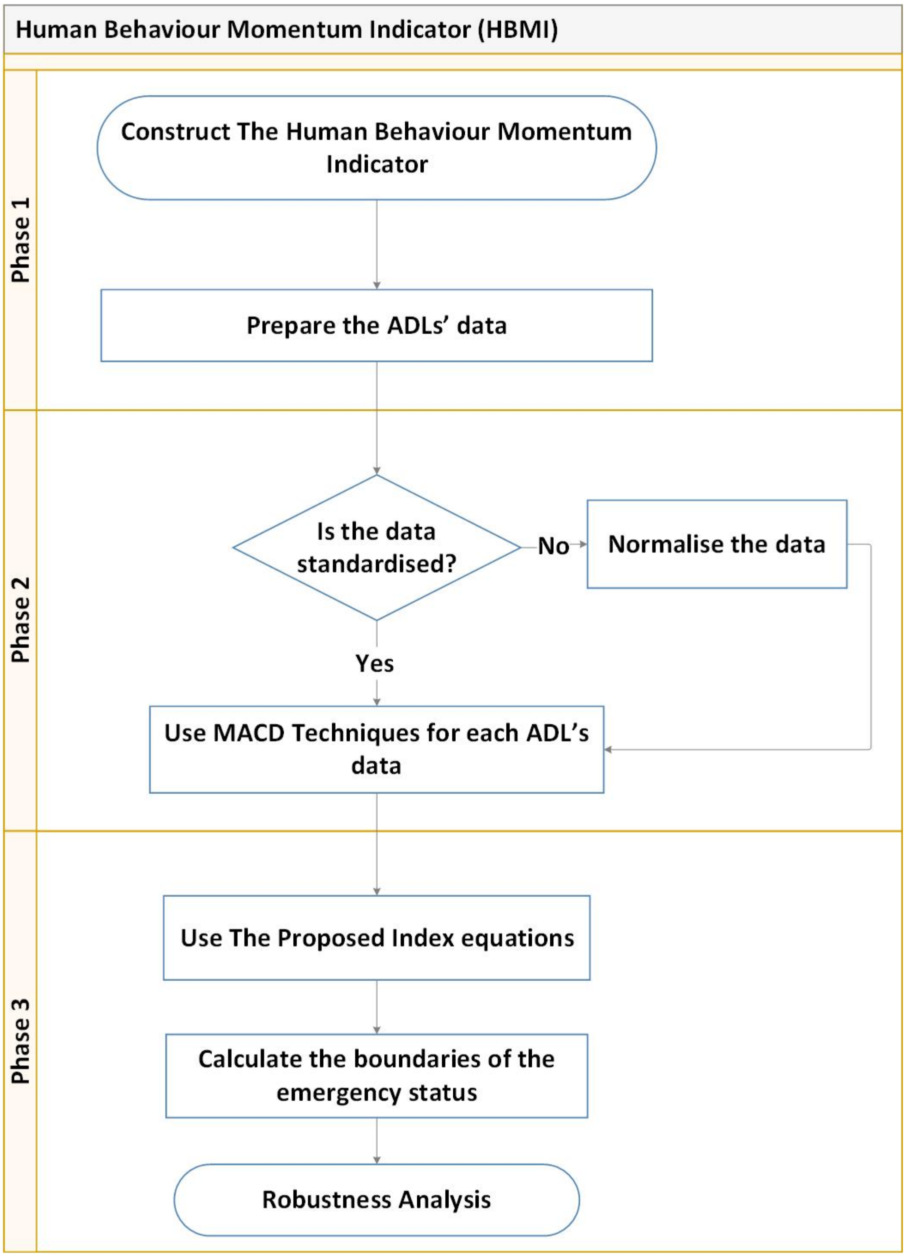

Figure 5.14: Human behaviour momentum indicator model.

- Calculate

$$avg = \sum Index/N. \tag{5.2}$$

where: N is the total number of data points

- Calculate
$$NB = Avg \pm 100. \tag{5.3}$$
where: NB is the Normal boundary.

- Calculate
$$WB = NB \pm AV. \tag{5.4}$$
where: NB is the Warning boundary and AV is the accepted value to represent the warning situation.

- Calculate
$$AB = WB \pm V. \tag{5.5}$$
where: AB is the Abnormal boundary and V is any value.

Equation 5.1 will compute the index values for each data point, Equation 5.2 is used to compute the average value of the index and Equations 5.3, 5.4 and 5.5 are used to calculate the three boundaries.

Figure 5.14 illustrates the proposed model. The technique is tested with different data sets and it shows superior results in terms of forecasting the activity and as well detected the abnormal values.

The HBMI is used to measure changes in the trend of the human behaviours' data and predict the direction of the trend. Moreover, it will present the situation of the human behaviour activity based on the idea of having three levels of the activity's situation which are Normal, Warning and Abnormal. The indicator will response to the changes that may happen in the data. On the other hand, it will indicate the changes after the MACD shows a sign of upcoming changes, which means the MACD will work as a part of HBMI's process. The HBMI is a special indicator technique, it presents together momentum and trends in data using MACD and it predicts the upcoming changes in the data. Furthermore, it shows the changes of the ADLs/ ADWs behaviour levels based on the indicator. This unique combination between these three signals could be applied to different kind of data charts such as daily, weekly or monthly. In this test several data sets are used to examine the HBMI.

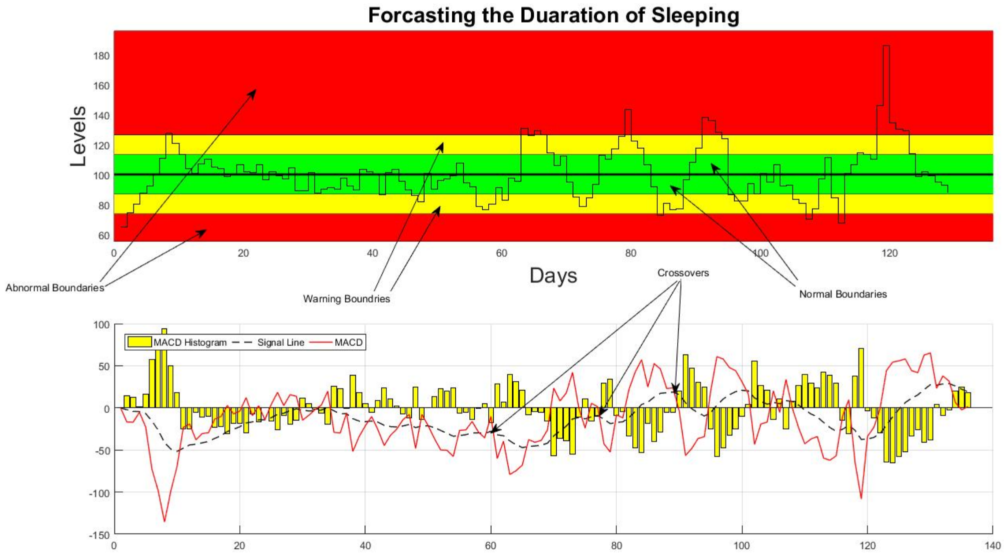

Figure 5.15: Using HBMI to forecast the duration of sleeping in a bedroom.

The following tips will interpret some of the results that may be seen in the graph:

- When the "MACD line" crosses the signal line that is called crossover, which is mean changes in data values will be identified.
- The direction of "MACD line" depends on the direction of the moving average cross. Therefore, when the "MACD line" direction is changed then the trend direction is changed too.
- When "MACD line" oscillates above the signal line indicates that the indicator signal in the top graph is decreased. Also, the indicator signal will be increased if the "MACD line" fluctuate below the signal line.
- In addition, if the "MACD line" accelerates sharply then the indicator will present sharp changes in the data. In other words, the increase or decrease of the indicator signals in the top graph means the changes of the trend's direction is happen in the same direction of the indicator.

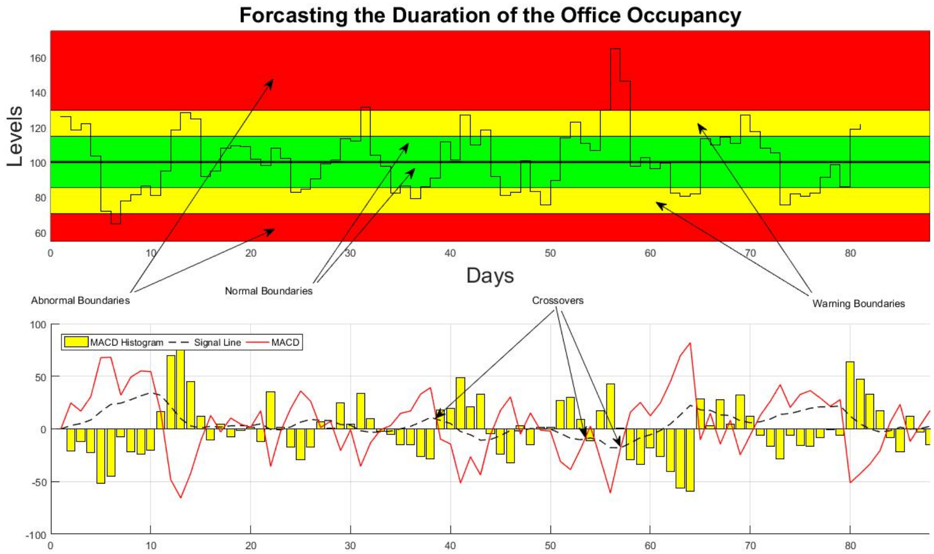

Figure 5.16: Using HBMI to forecast the duration of occupying an office.

In the first experiment, a data set of the duration of sleeping in the bedroom is used to test HBMI. The number of days (7, 28, 14) are used to calculate the MACD. Also, 7 days are used to calculate the indicator. The results show the HBMI technique is a very useful tool to detect trends and predict the changes that could happen in the trend based on historical data as well it is detects abnormal events based on its boundaries theory. For example, Figure 5.15 shows that the sign of upcoming change in the trend direction can be observed when the crossover occurs between the signal line and MACD line in the day 10. On the other hand, with the changes of trend's direction in the indicator the abnormal event can be detected too. For instance, the abnormal event detected in the day 118 when the MACD accelerates sharply from the day 117 to the day 118 and at same time crossover between the "MACD line" and signal line occurs between the days 117 and 118. Figure 5.15 shows an example of the results of forecasting a duration of sleeping in a bedroom. At the same time, the changes can be seen by looking to the data in Table 5.6, which shows the results of using HBMI to predict the direction of trend. The first column shows uptrend after day 10 and the second column shows downtrend after day 89.

The proposed technique (HBMI) is examined with another real data set, which is representing the duration of occupying a smart office. The numbers of days (5, 20, 10) are used to calculate MACD. Again, the results of using HBMI technique shows it is a very useful tool to detect trends and predict the changes that could happen in the trend based on historical data. For example, when the crossover occurs between the signal line and "MACD line" in the day 14 it can be seen that the upper side of Figure 5.16 the changes of the trend's direction. The trend direction is changed when the crossover occurs between the "MACD line" and the signal line in the day 54, moreover, the abnormal event is detected when the "MACD line" fluctuated sharply from day 54 to day 55. As it is shown in Figure 5.16 the HBMI can detect changes that may happen in ADWs. On the other hand, the changes can be seen by looking to the results data in Table 5.7, which shows the results of using HBMI to predict the direction of trend. The first column shows down-trend after day 10 and the second column shows up-trend after day 43.

On the other hand, HBMI can detect the abnormal events in the data sets with good accuracy and it can be used to predict the abnormal events as well when the "MACD line" is accelerates sharply between the data points. Tables 5.8 and 5.9 show examples of detecting abnormal events. Comparing data points in the second column of both tables clearly shows the abnormal events and these events can be seen in the red area in Figures 5.15 and 5.16.

Finally, the HBMI technique is tested with different data sets and it shows good results of the trend's direction prediction and detection of changes in ADLs/ ADWs.

5.3.4 Moving Average Convergence/Divergence

The MACD is special indicator technique because it presents together momentum and trend at the same time. This unique combine between trend and momentum could be applied to different kind of data charts such as daily, weekly or monthly. In this test three settings for MACD which are the difference between the (7, 28), (14, 28) and (14, 56) periods of EWMAs. The MACD Line is calculated based on the difference between two moving averages. This means MACD values are

Table 5.6: A sample of results of using HBMI with the duration of sleeping.

Data after Day 10 **represent total duration in minutes**	**Data after Day** 89 **represent total duration in minutes**
526.26	452.12
527.32	443.82
529.09	423.42
540.17	421.22
539.47	421.13

Table 5.7: A sample of results of using HBMI With the duration of occupying an office.

Data after Day 10 **represent total duration in minutes**	**Data after Day** 43 **represent total duration in minutes**
464.56	304.35
467.26	373.45
400.20	413.01
288.24	433.70
262.23	366.90

Table 5.8: A sample of using HBMI to detect abnormal events in the duration of sleeping in a bedroom.

HBMI low boundary	**Original Data Values**
80.69	341.58
70.13	655.20
77.11	323.35
HBMI High boundary	**Original Data Values**
125.42	415.55
143.15	187.11
122.54	300.03

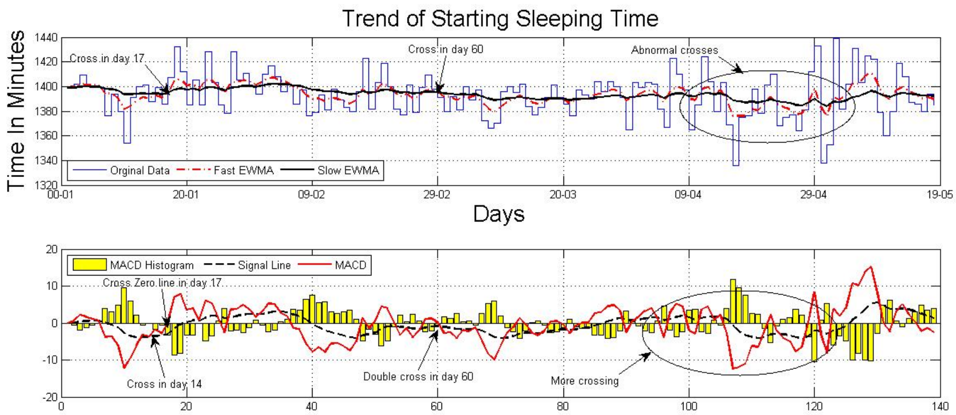

Figure 5.17: Using MACD with 7 and 28 data points.

dependent on the datum point of the underlying security.

The crossovers signal occurs when the fast EWMA has crossed the Slow EWMA and the MACD Line oscillates above and below the zero line. The direction of MACD depends on the direction of the moving average cross. Positive MACD indicates that the fast EWMA is above the slow EWMA. When the upside momentum is increasing, then positive values of MACD increase as the fast EWMA diverges further from the slow EWMA. Negative MACD values indicates that the fast EWMA is below the slow EWMA. Therefore, negative values increase when the fast EWMA diverges further below the slow EWMA. This means

Table 5.9: A sample of using HBMI to detect abnormal events in the duration of occupying an office.

HBMI low boundary	**Original Data Values**
60.47	479.63
54.90	285.76
74.58	522.36
HBMI High boundary	**Original Data Values**
115.26	457.58
142.17	189.11
100.59	443.06

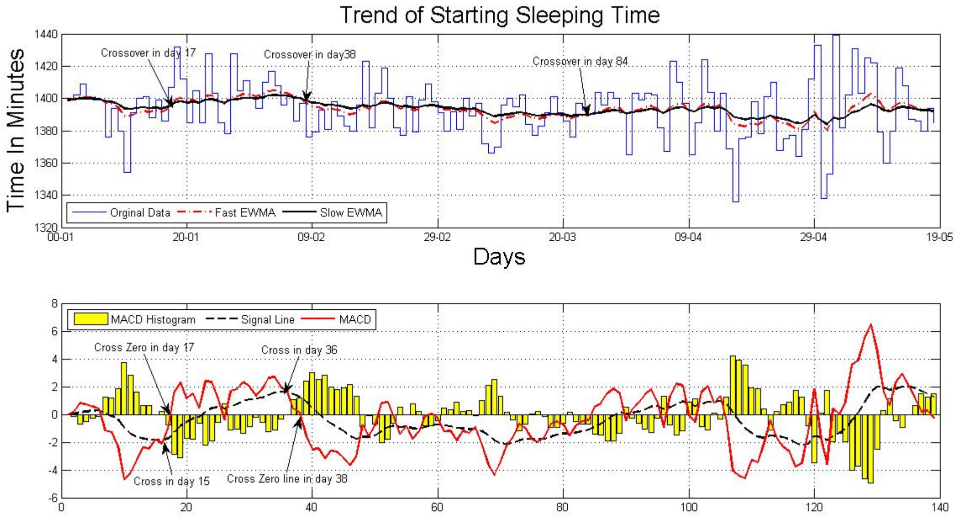

Figure 5.18: Using MACD 14 and 28 data points.

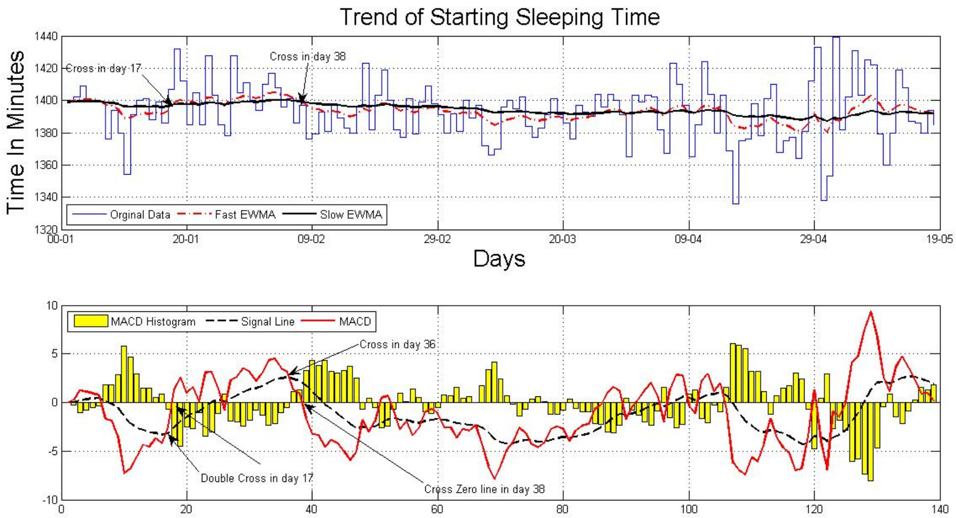

Figure 5.19: Using MACD 14 and 56 data points.

downside momentum is increasing.

In this experiment, the first pair of numbers (7, 14) are used in MACD algorithm to detect the trend and momentum for the data which are explained previously in this research. It is clear that, the MACD technique is very useful tool to detect trends and predict the changes that could be happen in the trend

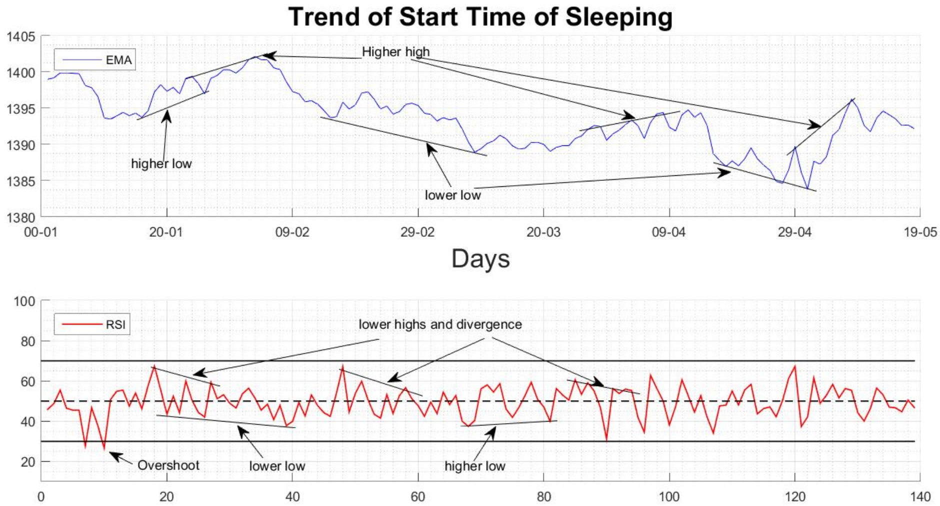

Figure 5.20: Using RSI with 7 and 28 data points.

based on historical data. For example, when the crossover occurs between the signal line and MACD line at the end of day 14 and crossover occurs between the MACD line and Zero line at the early of day 17 then crossover occurs between Fast EWMA and slow EWMA late in day 17, which show the change of sleeping start time trend, as it is shown in Figure 5.17. However, the same MACD algorithm is used with other pairs of numbers (14, 28) and (14, 56) and equivalent results are gained. Figures 5.18 and 5.19 are shown the results. The results show that, this technique could be used to detect the abnormal behaviour based on abnormal crossovers occur when using EWMA technique.

5.3.5 Relative Strength Index

The RSI tends to fluctuate between 40 and 90 is used to show the uptrend. It will act as support in technical analysis for the professional when it will be within the $40 - 50$ zones. These ranges may vary depending on RSI parameters, the strength of trend and volatility of the underlying security. Figure 5.20 shows 14 days RSI for the start time of sleeping during the period from December until April. There was one overshoot below 40 in January, but RSI held the $40 - 50$ zone several times. On the flip side, RSI tends to fluctuate between 10 and 60 to

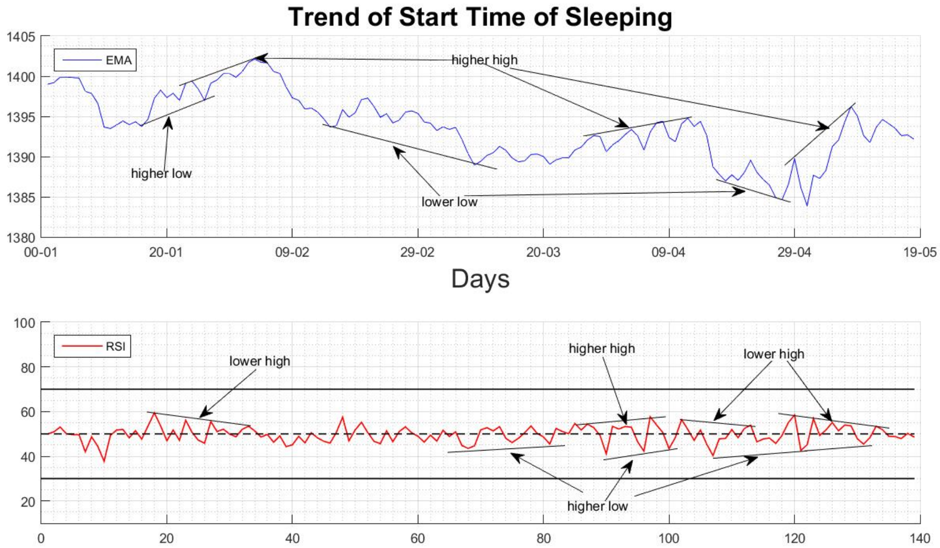

Figure 5.21: Using RSI 14 and 28 data points.

display the down-trend with the 50 − 60 zone acting as resistance. Figure 5.21 shows 7 days RSI for the same data that is used in this example. It is shown that the RSI with small numbers will be more sensitive to changes that may happen in data, which will make the interpretation of trends harder.

5.4 Discussion

In this research, trend analysis techniques to detect trends in human behaviour are investigated. Also, forecasting techniques that can be applied to ADLs and ADWs data are investigated. Our results demonstrate the feasibility of producing data interpretation to support the carer or managers with general information about his/her patient or worker. The conducted experiments introduce superior results from using trend analysis techniques to interpret the humans' activities. In addition, The conducted experimenters produced good results of using forecasting techniques that are working based on trend analysis techniques to forecast and predict the humans' activities and the direction of trends in human behaviour.

To use trend analysis techniques in this research, data preparation is the first step to overcome some issues such as dealing with missing data, converting data from binary to the suitable format that could be used later with trend analysis techniques and extracting the exact data for each activity. Therefore, a prototype of the data was created to make the data more understandable for the reader. The MA techniques are used to detect trends in studied data. Starting with SMA which shows the ability to display the trend and its direction as well. SMA could be used to find out the abnormal events that are happening in the data. The obtained results from applying CMA, MMA and SMM techniques show that; they can detect trends and trend direction but cannot be used to detect the abnormality in our data. The last technique of moving average family that is used in this research is EWMA. EWMA is the best technique in the moving average family that could be used to detect trends in this kind of studied data, because of its sensitivity to the changes in data and it has the ability to detect trends, trend direction and the abnormal events that have happened. SKT is used to detect trends in human behaviour. This technique displays good results in terms of seasonality and trends. The comparison between the results and indexed value shows that this technique could detect the trends in our data. However, this technique cannot be used to predict the future and it will not display the direction of trend.

The forecasting and prediction techniques used in this research that can be applied to ADLs and ADWs data are investigated. The proposed solutions demonstrate the feasibility of interpreting this kind of data. The indicator techniques are used to predict the future of the trend direction in our data sets. The HBMI and RSI techniques are used for this purpose in this research. Both of them present good results in terms of detecting trends and trend's direction. The HBMI shows more information about the trend and its orientation and the levels of the fluctuation of ADLs. The RSI sometimes needs to be used with another technique to support its results. AFMA is used to predict N data points along side is is used as trend analysis technique. It presents good results in terms of prediction. Moreover, in contrast to using EWMA which is the best technique in order to detect trends in our data and because of its sensitivity to the changes in data, it shows that it has the ability to identify the concealed abnormal events. ANFIS is used to predict new values of our data sets. It presents good results in terms of estimating new values of such data that we have. When using ANFIS with this kind of data, it is essential to smooth the data and optimise the number of membership functions.

Chapter 6

The Human Behaviour Indicator

6.1 Introduction

Progressive change in people's activities is a subject of research interest that has recently attracted more attention [175]. These could be activities of daily living (ADL) or activities of daily working (ADW). A number of medical conditions and their treatments are associated with activities disorders such as reduced movement over time. Parkinson's Disease (PD), for example, is characterised by slowness of movement [176], and some Mild Cognitive Impairment (MCI) causes a slight but noticeable and measurable decline in ADL [177, 178].

Understanding progressive changes in human behaviour is the key challenge in formulating an effective intelligent environment. It is essential to find a suitable technique that models human behaviour and interprets progressive changes in the participant's behaviour. Many researchers have investigated human activity recognition in order to solve the problem of activity extraction and prediction. These studies have used different methods and techniques to identify the behaviour of occupants based on temporal data gathered from sensor networks.

There are several methods have the potential to represent and model human behaviour because they are capable of representing random activities, dependencies and temporal variations within the data [179]. Probabilistic models such as the Hidden Markov Model (HMM) and Bayesian belief networks can be used to model human behaviour [179, 180, 181, 182, 183]. Soft computing and machine-learning techniques such as artificial neural networks (ANNs) and support vector machines (SVM) can also be used in this way [184, 185]. However, all mentioned techniques face difficulties in processing large amounts of low-level sensory data; therefore, it is essential to transform the sensory data into a suitable format that can be processed [186].

Trend analysis techniques are used to understand the participants' general health variation and behaviour evolution. Analysing the activities of a participant can only be fully understood if the daily activities are examined in terms of person, place and time. Numerous researchers have investigated trend analysis techniques that can be applied to human behaviour activities extracted from smart environments. Researchers on the BackHome [187] and iCarer [188] projects introduced monitoring systems to send feedback on any changes in the participant's behaviour and habits to the carer. In another effort, Forkan et. al. in [189] used trend analysis to detect trends in their data, but it is not clear which trend analysis technique of they used. Saives et. al in [190] presented a model to improve the autonomy of medically monitored behaviour changes for patients in a smart home.

To be able to have a measure of progressive changes, the first step is to integrate some pervasive measurements into home and work environments to collect the data representing ADL or ADW. There are already many off-the-shelf products available to collect such data. Once individual activities are recognised, identifying trends in users' behaviour over a period of time will provide useful information. Trends in people's behaviour are used to identify progressive changes and predict behavioural abnormalities.

To address the challenges of the research, this chapter introduces a composite indicator representing overall progressive changes in a behavioural pattern. This approach is novel and the proposed approach has not been used for behavioural analysis before, though synthetic composite indicators (SCI) have been proposed to interpret progressive changes of aggregated variables in a variety of economic and policy areas. For example, the consumer price index (CPI) measures changes in the prices of goods and services that households consume [191].

This Chapter is organised as follows: in Section 6.2; the new Human Behaviour Indicator and its constructions are introduced; three case studies using different data sets are discussed in Section 6.3 and the summary and suggestions for future work are presented in Section 6.4.

6.2 The Human Behaviour Indicator

To represent human behaviour evolution through a simple measure requires some method of combining several types of data. This will require the construction of a composite indicator which can be viewed as a means of reducing multivariable inputs into a single and meaningful output that can be interpreted by non-experts. After giving a brief overview of composite indicators below, the Human Behaviour Indicator (HBI) will be explained.

6.2.1 Composite Indicators

The aggregated output usually represents a holistic view using a single numerical score and/or an ordinal rank. To achieve a single output value, the index must go through accurate development steps. These steps involve consideration of included or excluded variables, handling missing data, the given weight for each variable etc.

There are two major types of aggregation technique: additive (linear) and multiplicative (geometric or nonlinear) [192]. Additive aggregation technique can offer compensability between aggregated indicators; if an indicator has poor performance, then an indicator that has a significant value can cover the poor one, which may result in the creation of a biased composite indicator.

6.2.2 The Construction of the Human Behaviour Indicator

The schematic diagram of the proposed framework to create the Human Behaviour Indicator (HBI) is shown in Figure 6.1, and comprises three phases:

- In the first phase, the ADL or ADW data is collected and pre-processed. This includes the handling of missing data and the extraction of the features that will be involved in representing the holistic view of the activity. Each

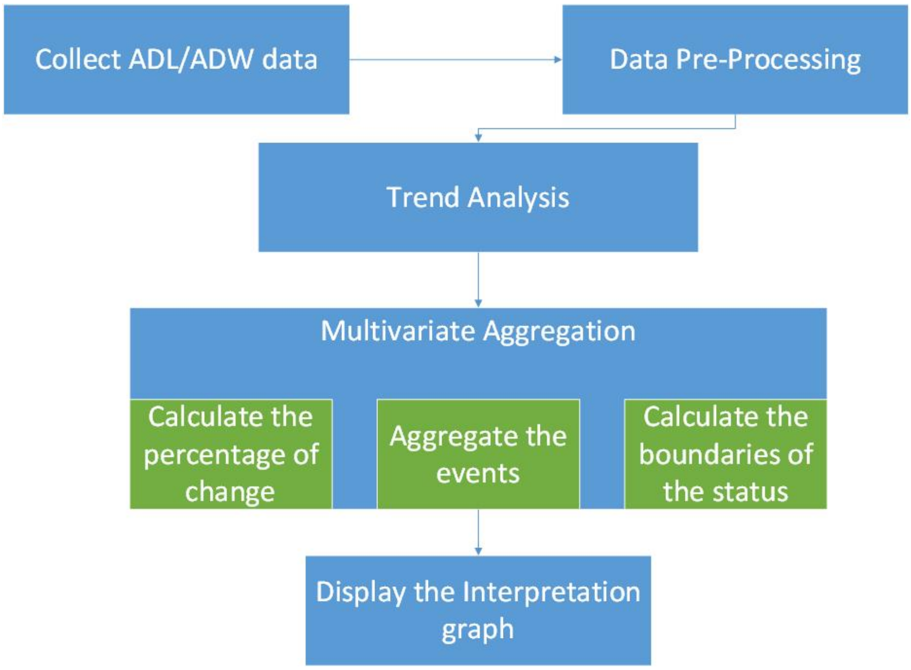

Figure 6.1: Proposed framework to create the Human Behaviour Indicator.

feature represents specific information from a single activity; for example, the start time and the duration of each event. For more details see Chapter 3.

- In the second phase of the proposed framework, the data are analysed using trend analysis techniques. For more details see Chapter 4.

- Finally, in the third phase of development, the data created from the trend analysis is standardised to a uniform unit of measurement and then aggregated. The boundaries of the three levels of the holistic view of the monitored person are also generated. These levels are the level of normal activity behaviour, the warning level to represent minor changes in activity behaviour, and the abnormal level to detect the outlier activity behaviour.

6.2.2.1 Multivariate Aggregation

To identify the overall changes and ultimately to forecast changes in the ADL or ADW, a new indicator, the Human Behaviour Indicator (HBI), is introduced.

This stage has two main tasks; the first task is to build the composite indicator, which will compute the progressive changes in the behaviour based on the events that are performed in the entire day. The events are the actions that the participant performs during his/her activity. Each event has a start time and a duration. For example, when the person works with a PC and sits on the chair, the start time of the activity will be computed. At the same time, the start time of using the chair until he/she leaves it as an event will be computed. This idea of using the events gave us the opportunity to work with overlapped events and activities. Therefore, we were able to compare or measure the changes of events on different bases such as daily or weekly. In this research, the percentage of changes are used between the events when compared to each other as the best way to measure the changes. However, this task had three steps, as follows:

- There are diverse ways in which human behaviour events could be aggregated and to calculate the percentage of change of the compared events. As part of our investigation, it is found that Weighted Human Behaviour Indicator (WHBI) expressed in Equation 6.1 and the Exponential Weighted Human Behaviour Indicator (EWHBI) expressed in Equation 6.2 are the best candidates for data represented in this chapter. Each equation has its way to reach the final results but both of them gave the same final results. Therefore, either of them could be used as a start point of the calculations.

- Calculate

$$Pr_i = \frac{\sum (P_{i+n}).2}{\sum (P_i).(P_{i+n})}. \tag{6.1}$$

- Calculate

$$Pr_i = \frac{(\alpha) \sum (P_{i+n}).2}{(1-\alpha) \sum (P_i).(P_{i+n})}. \tag{6.2}$$

  where: P is current data point value after using EWMA (i.e. start time, duration) of the activity for i=1,2,3,.. and n is the number of data points used to calculate EWMA.

- Calculate the summation of changes in events based on all daily events.

$$total_d = \sum Pr_i. \tag{6.3}$$

where: d is the number of each day and i=1,2,3,.. represents the percentage of change of each compared event of that day.

- Calculate the percentage of total changes of all daily events.

$$Index_i = (total_d/total_{d-n}).100. \tag{6.4}$$

where: i,d=1,2,3,.. and n is the number of data points that are used to calculate EWMA.

The second task is to calculate the boundaries of the changes. In other words, the index is centralised to 100, because the indicator will present the percentage of the changes in the overall of daily activities after calculating the percentage of change between the values of compared events' features (i.e start time and duration of each event). The system has three boundaries on each side of the centre value (100). These boundaries represent three main categories: normal, warning and abnormal. The following steps show the process of calculating the indicator and the boundaries:

- Calculate the average of all Indexes' values.

$$avg = \sum Index/N. \tag{6.5}$$

where: N is the total number of data points.

- Calculate the normal boundary.

$$NB = 100 \pm Avg. \tag{6.6}$$

where: NB is the normal boundary.

- Calculate the warning boundary.

$$WB = AV \pm NB. \tag{6.7}$$

where: WB is the warning boundary and AV is the accepted value to represent the warning situation.

- Calculate the abnormal boundary.

$$AB = WB \pm P_t. \tag{6.8}$$

where: AB is the abnormal boundary and P_t is a data value at time t.

6.3 Case Studies and Discussion

In this section a brief description of data used in this research is provided before the proposed method of formulating the indicator is introduced. The data sets used for this research are collected from within smart environments equipped with some sensors to measure indoor activities of an occupant. The indoor movements are represented by the sequences of movement from one place to another. Monitoring the movements of a person occupying the environment and collecting the data representing his/her mobility can detect this person's transitions.

Table 3.3 shows the format of the data summarisation, which represents two people, each occupying a smart environment. The first person occupies a smart home and includes examples of his/her activates based on his/her sequential movements during the occupancy of a specific location. Each behaviour can be represented in two order levels. The first level is the activity sequence, which shows the person's movements. The second level is the sequence of actions, which shows the occupancy time of the activity that the person has performed. For example, walking to the kitchen is a moving activity and sitting in the living room is an occupancy activity.

Activities in a smart home are presented by five main categories: "Bedroom", "Living Room", "Kitchen", "Bathroom" and "Hall" activities, representing sleeping, socialising, eating, cleaning and moving between places, respectively. Activities in a smart office are presented by four main categories: "Chair", "PC", "PIR" and "Door" activities, representing sitting, working, occupancy and going in/out, respectively.

In this section, three separate case studies are reported. Two of these case studies have real data sets collected from smart environments. The first data set represent the ADL of an elderly person who lives in a smart home and the second

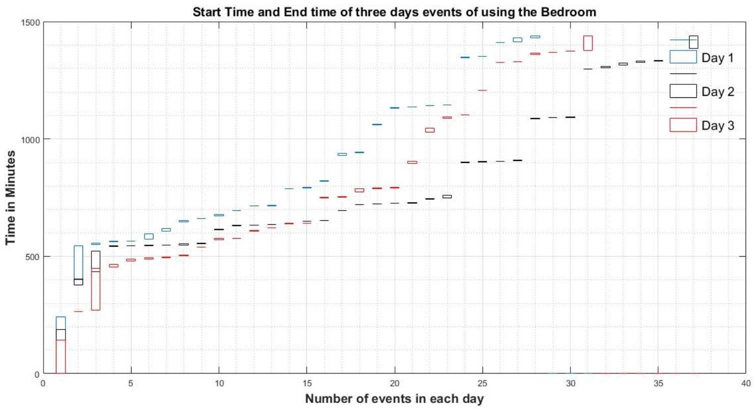

Figure 6.2: Sample of events in the Bedroom over three days.

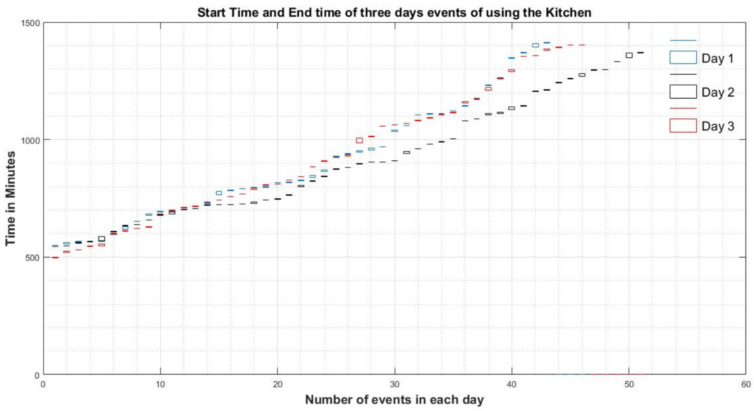

Figure 6.3: Sample of events in the Kitchen over three days.

data set represents the ADW of a person who works in a smart office. In the third case study, a simulated data set is used; this data set simulates an elderly person who lives in a smart home. Real data were collected using wireless motion sensors, PIR and door sensors. The output values of these sensors are discrete. In the smart home these values represent the occupancy of one area at a time, whereas in the smart office, the values may overlap because of the possibility of using different equipment at the same time; for example, the participant may use the PC while sitting on the chair and occupying the desk. However, the results and discussion presented below are based on one participant in each smart environment.

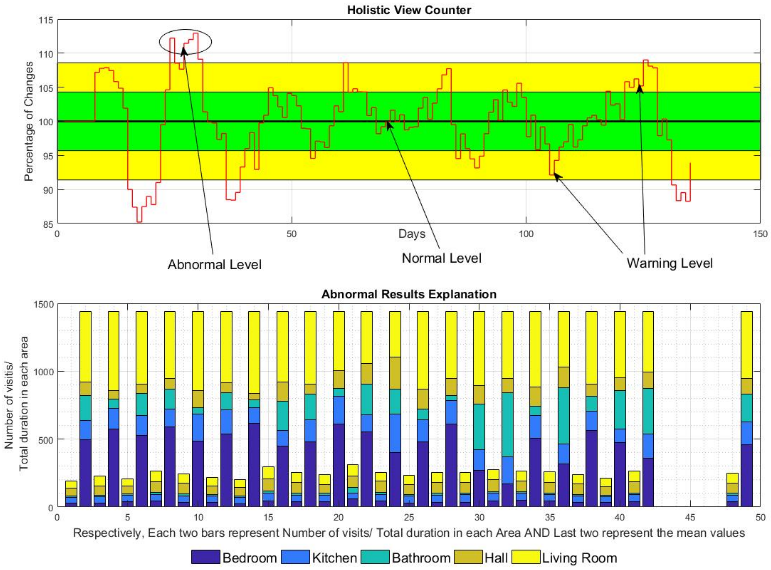

Figure 6.4: Holistic view of an elderly person's behaviour in a smart home.

6.3.1 Case Study 1: Combine all events of ADLs extracted from real data of a monitored person who lives in a smart home

This study aims to establish whether the proposed technique can show the overview of a person's status and detect abnormality within the behaviour of a participant who lives in a real environment. In this case study, the data represents all events of ADLs that are completed in different areas of a smart house. To demonstrate the holistic view of the monitored person's activity status based on the sensor network measurement, the data sets are represented in separate groups. Each group shows the start time and duration of each event in one area. These events represent the occupancy activities in each area of the smart home.

For example, Figure 6.2 and Figure 6.3 show the start time and duration time of events that represent the occupancy activities for the bedroom and kitchen, respectively. In both figures, a sample of only three days is depicted. These days are the same weekday of three weeks. It is very difficult to see each area's events in one graph if we are hoping to understand the overall behaviour changes or to interpret the behaviour; however the proposed technique can display a holistic view of the person's behaviour in one graph. For instance, Figure 6.4 shows the status of a person based on the collected data for 140 days.

As can be seen from Figure 6.4, the participant is often in a good situation. Frequently, the percentage of changes in this person's behaviour are around the average of his/her daily routine. However, on some days there are some abnormalities in the percentage of changes, which goes beyond the second boundaries that represent the red area in this graph. To explain what led to this result, the lower graph in Figure 6.4 shows the number of times the participant visited the areas and the total duration spent in each area, alongside the mean of number of visits and the total duration for each area. We thus have at least an idea of the differences in the patterns of normal and abnormal behaviour, as result of using our technique to interpret the overview of the participant's behaviour in an appropriate graph of all events that occurred in all areas in the home.

In addition, this technique makes it possible to identify the abnormality from low-level sensory signals even if we do not have detailed knowledge of the subject. Table 6.1 shows a sample of the real data recorded in a smart home, which represents the three level that are used in our technique. This information shows that our technique can measure the changes of human behaviour.

Table 6.1: A sample of the results of using HBI with the smart home's data.

Places	**Normal**		**Warning**		**Abnormal**	
	T.D	N.V.P	T.D	N.V.P	T.D	N.V.P
Bedroom	561	34	488	35	300	30
Kitchen	203	51	217	56	175	55
Bathroom	97	15	154	11	278	13
Hall	102	69	121	82	113	74
Living Room	477	56	458	72	572	73

T.D: Total duration, N.V.P: Number of visiting a place

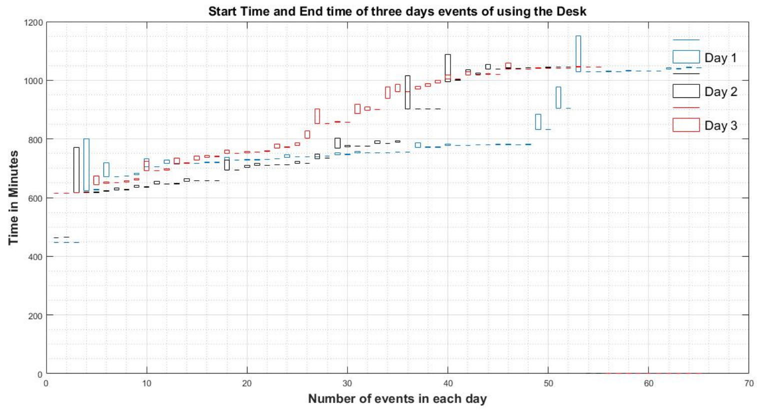

Figure 6.5: Sample of events using the desk over three days.

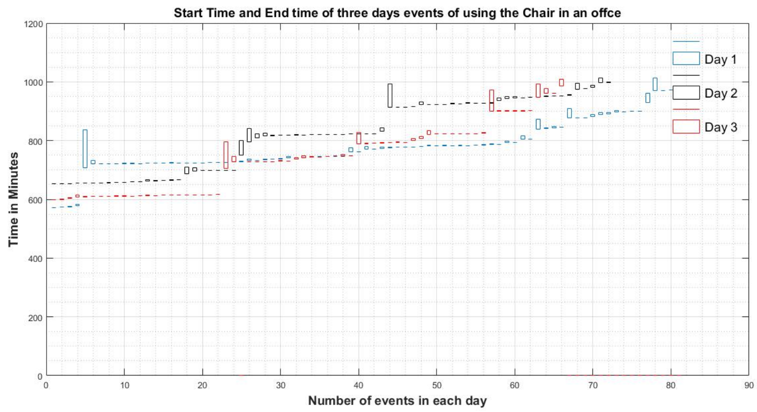

Figure 6.6: Sample of events using the chair over three days.

6.3.2 Case Study 2: Combine all events of ADWs extracted from real data of a monitored person who works in a smart office

The aim of this case study is to investigate whether the proposed technique can show an overview of the person's status and detect abnormality within the behaviour of a participant who works in a real environment. In this case study, the data represents all events of ADWs that are performed in different areas in a smart office. To demonstrate the holistic view of the monitored person's activity status based on the sensor network measurements, the data set is represented in

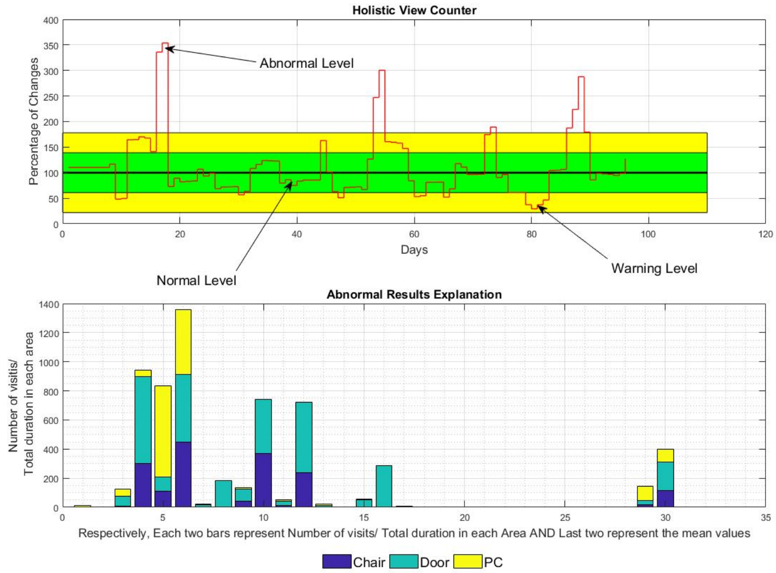

Figure 6.7: Holistic view of a person's behaviour in a smart office.

separate groups. Each group shows the start time and duration of each event in one area. These events represent the occupancy activities in each area in the smart office.

It is very important to notice that the data sets overlapped in their times. For example, when the person is using his/her PC, the PC's sensor will trigger; at the same time, the chair's sensor will trigger. Figure 6.5 and Figure 6.6 show the start times and durations of events that represent the occupancy activities for the desk and the chair, respectively. In both figures, only a sample of three days is depicted. These days are the same weekday in three weeks (e.g. Mondays). Applying our technique to such data can help to understand the overall behaviour changes or the behaviour interpretation by displaying the holistic view of the person's behaviour in one graph. For instance, Figure 6.7 shows the status of a person based on the collected data for three months.

As shown in Figure 6.7, the participant is using his/her office normally. However, on some days an abnormality occurs and we can see the change in the percentage of changes levels, which goes beyond the second boundaries that represent the red area in this graph. To explain what led to this result, the lower graph in Figure 6.7 shows the number of times the participants used the equip-

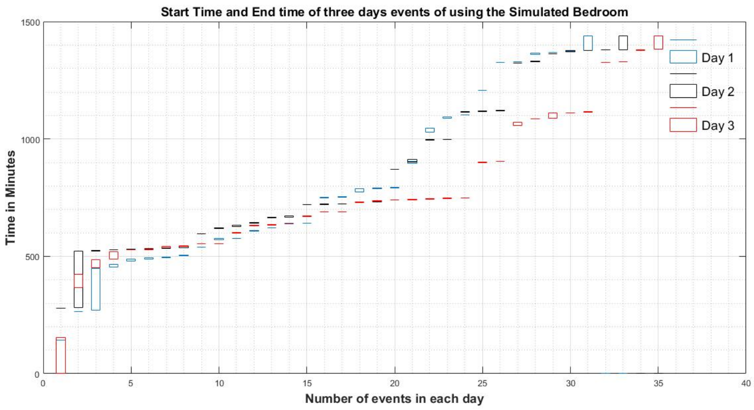

Figure 6.8: Sample of simulated events in the Bedroom over three days.

ment in the office and the total duration of each use, alongside the mean number of usage and the total duration of use of each piece of equipment. Therefore, we have at least an idea of the differences in the patterns in the normal and abnormal behaviour.

Table 6.2 shows that we gained important results from using our technique with data from multiple sensors that sometimes overlapped. The HBI shows the progressive changes in the participant's behaviour and interprets it in one relevant graph of all events using all sensors in the office. In addition, this technique makes it possible to identify the abnormality from low-level sensory signals, even if we do not have detailed knowledge of the subject.

Table 6.2: A sample of the results of using HBI with the smart office's data.

Places	**Normal**		**Warning**		**Abnormal**	
	T.D	N.T.S	T.D	N.T.S	T.D	N.T.S
Chair	401	83	338	54	303	9
Duration	404	56	513	60	595	67
PC	401	449	291	213	44	48

T.D: Total duration, N.V.P: Number of triggering a sensor

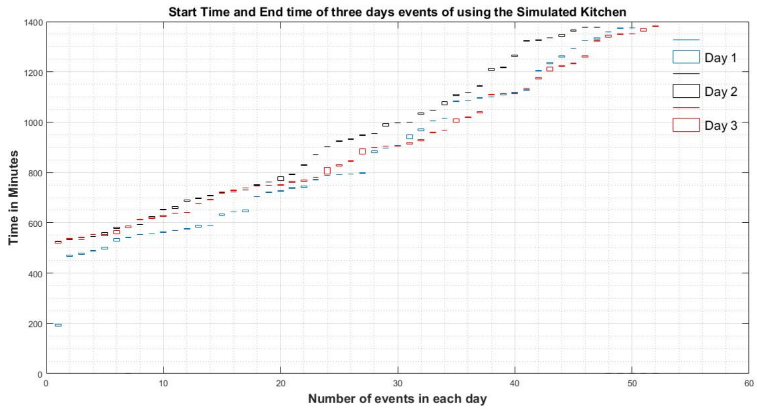

Figure 6.9: Sample of simulated events in the Kitchen over three days.

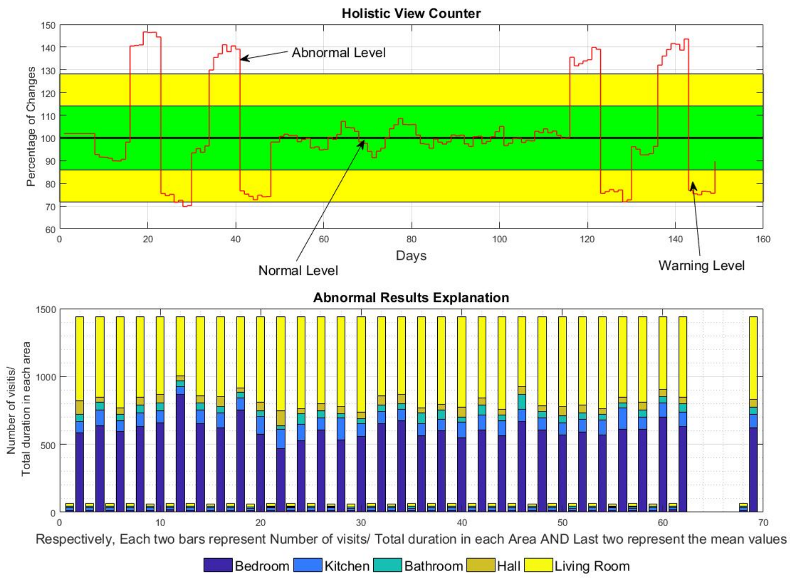

Figure 6.10: Holistic view of simulated activities of an elderly person in a smart home.

6.3.3 Case Study 3: Combine all events of ADLs extracted from simulated data of an elderly person

In this case study, simulated data prepared for this research to produce data that can be compared with real data, also it is used to demonstrate that the proposed technique can detect abnormalities in simulated data, as well to show the overview of the person's status. In this case study, the data represents all events of a less mobile person's ADLs that are simulated to be carried out in different areas in a smart house. The simulated data are grouped into events to demonstrate the holistic view of the monitored person's activities.

Each group shows the start time and duration of each event in one area. These events representing the participant's activities in each area in the simulated smart home. For example, Figure 6.8 and Figure 6.9 respectively show the start time and duration time of events that represent the occupancy activities for the bedroom and the kitchen. In both figures, a sample of only three days is depicted. These days are the same weekday in three weeks (e.g. Mondays). However, it is very difficult to understand the overall situation of the monitored person by observing each area's events graph, but we can easily understand the overall behaviour changes or the behaviour interpretation if we have a single graph to show the changes. For instance, Figure 6.10 shows the progressive changes of a person's behaviour based on the simulated data for around 150 days.

As shown in from Figure 6.10, the participant has some abnormal days based on the percentage of changes in this person's behaviour. On the other hand, we can see that this person has a more stable situation because the frequency of the percentage of changes in this person's behaviour are around the average of his/her daily routine. To explain what led to this result, the lower graph in Figure 6.10 shows the number of visits to the areas and the total duration in each area, alongside the mean of the number of visits and the total duration of each area used as a scale of changes. Therefore, we have at least an idea of the

Table 6.3: A sample of the results of using HBI with the simulated smart home's data.

Places	**Normal**		**Warning**		**Abnormal**	
	T.D	N.V.P	T.D	N.V.P	T.D	N.V.P
Bedroom	640	13	673	14	701	14
Kitchen	96	19	102	18	104	18
Bathroom	64	10	48	9	48	10
Go Out	79	1	65	1	49	1
Living Room	560	21	550	19	537	22

T.D: Total duration, N.V.P: Number of visiting a place differences in the patterns of normal and abnormal behaviour. Again, in this case study, we can see from the results from using our technique, that it can be used to interpret the overview of the participant's behaviour in an appropriate graph of all events that occurred in all areas in the home.

In addition, this technique makes it possible to identify the human behaviour levels that are proposed in our technique (normal, warning and abnormal) from low-level sensory signals as it is shown in Table Table 6.3.

6.4 Discussion

In this chapter, a novel technique for measuring progressive changes in a person's ADL/ADW behaviour is introduced that can deal with multiple activities and present the output in a single graph. The start time, duration and number of repeated events are the most important features in representing a large sensor data set. These features can also be used in the trend analysis to identify abnormal behaviour.

Furthermore, the results presented in this chapter show that HBI is a very promising approach to interpret binary data sets collected from smart environments. The data sets investigated in this chapter were based on single participants who occupied smart environments equipped with appropriate sensors. These sensors were used to record the events of ADLs and ADWs that represent the behaviour of the participant and monitor his/her progressive changes. The results presented in this work show that low-level sensors can provide valuable information about the patterns and behaviour of the participant. The proposed approach can be applied to smart environments.

This study aimed to investigate the possibility of predicting events and behaviours in our future work. The approach presented in this chapter can work with overlapping data sets in the office, but it still needs to be examined in terms of multiple occupancy.

Chapter 7

Conclusions and Future Works

7.1 Thesis Summary

This thesis presents a novel attempt to answer the research question from the practical and theoretical point of view. The results obtained from the used techniques in this research allow the conclusion that a smart environment equipped with sensory devices can provide understandable interpretation of the data sets collected from the smart environment. Additionally, they can provide information about the human behaviour evolution of the participant. The proposed approaches give better results when the monitored person has activities based on a daily routine.

The main aim of the research is to investigate effective techniques that can measure progressive changes in human behaviour using a sensor network to collect the data from smart environments. The investigated techniques are aimed to support independent living or working with people who use a smart environment to monitor their behaviour and identify their ADL or ADW. The research is conducted to enhance the efforts for better understanding of human behaviour evolution by identifying trends in the data sets that represent their behaviour and to predict the direction of the trends within these data sets.

The provided data sets for this investigation are collected from real environments and from a simulator developed for this research. The simulator is built with the ability to include trends in the ADL. The real data sets represent the occupancy of a smart home and smart office. Both are used to extract the ADL and ADW patterns of the participants. The data sets are mainly long sequences of binary data produced by sensor networks, so different approaches for data visualisation and compression are investigated. In addition, different techniques are investigated to identify abnormal behaviour in ADLs or ADWs. The new measures known as HBMI and HBI are used to measure the human behaviour evolution in the patterns of participants and to predict the direction of trends in the patterns of the behaviour.

In summary, throughout this research, original knowledge is obtained to understand the binary data sets' visualisation, identification and prediction. The research conclusions with critical discussion and the direction of some future works are presented in the remaining sections of this chapter.

7.2 Concluding Remarks

This thesis attempts to provide an analysis of identification and prediction of trends and their directions for human behaviour evolution in ADL and ADW. Conclusions for various aspects of the project are presented below.

7.2.1 Activity Representation and Visualisation

In this thesis, the need for flexible and efficient techniques to visualise and represent large binary data sets are highlighted. Data representation techniques are used and evaluated using data sets obtained from the simulator and smart environments. For example, the start-time and duration of events are successfully used to convert and represent the binary sensor data sets, which are used to measure the changes in human behaviour. In addition, by using the start-time and duration of all events, the abnormalities in the participant's behaviour are detected and the large data sets are clearly visualised. The results show that these techniques can convert the original long series of binary data sets to time series data sets without losing their original values and information.

7.2.2 Trends in Human Behaviour

In this thesis, the relationships between frequent behaviour patterns of the participants which represent ADL or ADW in smart environments are studied. We are interested in identifying trends in human behaviour and predicting the changes in the trend's direction. The understanding of the progressive changes in the participant's behaviour based on trend analysis is a result of this research.

This study particularly investigated trend analysis in human behaviour, using indicators to measure the direction of detected trends in binary data sets which represent the daily events in smart environments. The investigation in this research is aimed to assess techniques that can identify the trends and their changes of directions in the pattern of human behaviour. Additionally, it aims to distinguish between the normal and abnormal behaviour patterns of participants in smart environments.

Different techniques are applied to find the trends within the data sets that represent the daily events. These techniques include SMA, CMA, MMA, SMM, SKT and EWMA. Each of these techniques has a potential to identify trends in data sets representing human behaviour activities. However, the results of this investigation show that the EWMA technique is better than other trend analysis techniques for our binary data sets. EWMA can detect the changes in data quickly. However, as the data sets in human behaviour could change many times in different directions, it is rather difficult to predict the trend's direction within the sensor data.

To tackle the prediction of trend direction in the data and to identify abnormalities in users' behavioural patterns, the new measure HBMI is proposed. HBMI can predict upcoming changes in the behavioural data, benefiting in that from using MACD. The results gained from HBMI are compared to RSI indicator results. The HBMI gives better results in terms of detecting abnormalities and the prediction of upcoming changes in behaviour. The APMA is proposed to predict behavioural data values. APMA is built based on EWMA.

In addition, ANFIS is used in combination with EWMA. The events' data will be smoothed by EWMA then ANFIS will predict future activities data. The proposed techniques are evaluated using several data sets to identify and predict the trends in human behaviour patterns. Additionally, based on the gained results and knowledge from this research, these techniques could be used in other domains such as monitoring security and safety in smart environments.

7.3 Human Behaviour Evolution Extraction and Measurement

This research shows that the pattern of human behaviour evolution can be extracted and predicted in smart environments. We are mainly interested in identifying similarities and measuring the changes in daily events and understanding progressive changes in human behaviour. The understanding of the participant's behaviour evolution is a result of this research.

This study has particularly investigated measurement of the similarity in binary data sets which represent daily events in smart environments. The investigation in this research is aimed to assess measurements that can identify the changes in the pattern of human behaviour, and to distinguish between normal and abnormal behaviour patterns of participants in smart environments. The HBI technique is proposed and applied to binary time series to measure progressive changes in human behaviour and to detect abnormal behaviour. The HBI allows the supervisor/carer to observe any changes to patterns on a daily basis and it can interpret and explain the observed changes. The results from HBI are tested and applied to real data sets in two case studies and they show their ability to measure the changes and explain them.

In general, the major findings of this research in terms of measuring the changes in human behaviour activities using binary time series data, collected from smart environments, are listed below:

1. Very limited research is reported in the literature on measurement of changes in human behaviour and analysis of binary time series, in particular for analysing real binary data sets collected from smart environments. This research has helped with better interpretation and understanding of binary time series using the investigated techniques.

2. Real data sets collected from sensors in smart environments could lead to a large amount of data. This binary data is hard to work with and needs to be converted into time series using proper techniques that must not be misleading about the results.

3. Using EWMA has helped to be able to detect trends in the participant's behaviour and predict the future of his/her behaviour.

4. The demonstrated results indicate that the combination of prediction and trend analysis techniques produced a better understanding of time series. For example, HBMI gives better results than using RSI in terms of understanding the trends and observing the changes in the participant's behaviour. Additionally, the results presented in this research show that APMA and the combination of EWMA and ANFIS are very promising approaches for binary data sets collected from smart environments.

5. The importance of the extraction of daily events from the raw data (time sequence) as inputs to the used techniques is investigated.

6. The impact of the aggregation of daily events and observation of the behaviour changes using HBI technique is investigated. This technique gives clear results of behavioural changes and interprets them in a suitable graph. The HBI gives superior results even when data are overlapped such as data collected from a smart office.

7. Overall, this research could play a vital role in the fields of observing human behaviour changes with more development.

7.4 Directions for Future Works

Further investigation, in which future works could proceed, is suggested below:

- A further direction following this is to investigate the implementation of the approaches presented in this thesis for multiple users in smart environments. The approaches are not tested using data sets representing multiple occupancies i.e. when the elderly person has a companion pet, which is true in some cases, or having more than one person occupying the smart environment. The complex task is to distinguish between the participant's normal pattern and the situation when he/she has visitors at a specific time.

- This study will serve as a base for future studies in which a combination of data sets from several sensors can be used. The current study has only examined human behaviour changes in an individual user of a smart environment using binary data sets. A future study using data gathered from other sensors (e.g. non-binary) that may be found in smart environments to provide big data sets for the experiments would be very interesting. In addition, using data sets from sensors that can handle multi-attribute vectors could give more interpretation and explanation of these data sets.

- Building a friendly user interface system to train and test these techniques online would be interesting.

- The progressive changes techniques used in this research could be extended to measure more complex human behaviour.

References

[1] United Nations. Department of Economic, *World population ageing 2009*, vol. 295, United Nations Publications, 2010. 1

[2] Mukhtiar Memon, Stefan Rahr Wagner, Christian Fischer Pedersen, Femina Hassan Aysha Beevi, and Finn Overgaard Hansen, "Ambient assisted living healthcare frameworks, platforms, standards, and quality attributes", *Sensors*, vol. 14, no. 3, pp. 4312–4341, 2014. 1

[3] Ashley E Nixon, Joseph J Mazzola, Jeremy Bauer, Jeremy R Krueger, and Paul E Spector, "Can work make you sick? a meta-analysis of the relationships between job stressors and physical symptoms", *Work & Stress*, vol. 25, no. 1, pp. 1–22, 2011. 1

[4] Daniel C Ganster and Christopher C Rosen, "Work stress and employee health: A multidisciplinary review", *Journal of Management*, vol. 39, no. 5, pp. 1085–1122, 2013. 1

[5] Ramakrishna Ayyagari, Varun Grover, and Russell Purvis, "Technostress: technological antecedents and implications", *MIS quarterly*, vol. 35, no. 4, pp. 831–858, 2011. 1

[6] Narayanan C Krishnan and Diane J Cook, "Activity recognition on streaming sensor data", *Pervasive and Mobile Computing*, vol. 10, pp. 138–154, 2014. 1, 3, 10

[7] Norman Wolkove, Osama Elkholy, Marc Baltzan, and Mark Palayew, "Sleep and aging: 1. sleep disorders commonly found in older people",

Canadian Medical Association Journal, vol. 176, no. 9, pp. 1299–1304, 2007. 1

[8] K Manabe, T Matsui, M Yamaya, T Sato-Nakagawa, N Okamura, H Arai, and H Sasaki, "Sleep patterns and mortality among elderly patients in a geriatric hospital", *Gerontology*, vol. 46, no. 6, pp. 318–322, 2000. 1

[9] Parisa Rashidi, Diane J Cook, Lawrence B Holder, and Maureen Schmitter-Edgecombe, "Discovering activities to recognize and track in a smart environment", *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 4, pp. 527–539, 2011. 2

[10] Juan C Augusto, Vic Callaghan, Diane Cook, Achilles Kameas, and Ichiro Satoh, "Intelligent environments: a manifesto", *Human-Centric Computing and Information Sciences*, vol. 3, no. 1, pp. 12, 2013. 2

[11] G. Virone, B. Lefebvre, N. Noury, and J. Demongeot, "Modeling and computer simulation of physiological rhythms and behaviors at home for data fusion programs in a telecare system", in *5th International Workshop on Enterprise Networking and Computing in Healthcare Industry*, 2003, pp. 111–117, ID: 1. 3

[12] Husheng Li and Zhu Han, "Catch me if you can: An abnormality detection approach for collaborative spectrum sensing in cognitive radio networks", *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3554–3565, 2010. 3

[13] Bin Xie and Qing Wu, "Hmm-based tri-training algorithm in human activity recognition with smartphone", in *IEEE 2nd International Conference on Cloud Computing and Intelligent Systems (CCIS)*, 2012, vol. 01, pp. 109–113, ID: 1. 3

[14] Chin-De Liu, Yi-Nung Chung, and Pau-Choo Chung, "An interaction-embedded hmm framework for human behavior understanding: with nursing environments as examples", *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 5, pp. 1236–1246, 2010. 3

[15] Quentin Bramas and Sébastien Tixeuil, "The complexity of data aggregation in static and dynamic wireless sensor networks", *Information and Computation*, vol. 255, pp. 369–383, 2017. 3

[16] Ahmad Lotfi, Caroline Langensiepen, Sawsan M Mahmoud, and Mohammad Javad Akhlaghinia, "Smart homes for the elderly dementia sufferers: identification and prediction of abnormal behaviour", *Journal of Ambient Intelligence and Humanized Computing*, vol. 3, no. 3, pp. 205–218, 2012. 4

[17] Christian Debes, Andreas Merentitis, Sergey Sukhanov, Maria Niessen, Nikolaos Frangiadakis, and Alexander Bauer, "Monitoring activities of daily living in smart homes: Understanding human behavior", *IEEE Signal Processing Magazine*, vol. 33, no. 2, pp. 81–94, 2016. 4

[18] Roberto Casas, Rubén Blasco Marín, Alexia Robinet, Armando Roy Delgado, Armando Roy Yarza, John Mcginn, Richard Picking, and Vic Grout, "User modelling in ambient intelligence for elderly and disabled people", in *International Conference on Computers for Handicapped Persons*. Springer, 2008, pp. 114–122. 5

[19] 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, 2013. 5

[20] Alvaro Muro-De-La-Herran, Begonya Garcia-Zapirain, and Amaia Mendez-Zorrilla, "Gait analysis methods: An overview of wearable and non-wearable systems, highlighting clinical applications", *Sensors*, vol. 14, no. 2, pp. 3362–3394, 2014. 5

[21] Qin Ni, Ana Belén García Hernando, and Iván Pau de la Cruz, "The elderlys independent living in smart homes: A characterization of activities and sensing infrastructure survey to facilitate services development", *Sensors*, vol. 15, no. 5, pp. 11312–11362, 2015. 9

[22] Oscar D Lara and Miguel A Labrador, "A survey on human activity recognition using wearable sensors.", *IEEE Communications Surveys and Tutorials*, vol. 15, no. 3, pp. 1192–1209, 2013. 9

[23] Roberto Casas, Rubén Blasco Marín, Alexia Robinet, Armando Roy Delgado, Armando Roy Yarza, John Mcginn, Richard Picking, and Vic Grout, "User modelling in ambient intelligence for elderly and disabled people", in *International Conference on Computers for Handicapped Persons*. Springer, 2008, pp. 114–122. 9

[24] Wadha Al-Marri, Amin Al-Habaibeh, and Hafez Abdo, "Exploring the relationship between energy cost and people's consumption behaviour", *Energy Procedia*, vol. 105, pp. 3464–3470, 2017. 9

[25] Ahmed Nabil Belbachir, Mario Drobics, and Walter Marschitz, "Ambient assisted living for ageing wellan overview", *e & i Elektrotechnik und Informationstechnik*, vol. 127, no. 7-8, pp. 200–205, 2010. 10

[26] Sawsan M. Mahmoud, M. Javad Akhlaghinia, Ahmad Lotfi, and Caroline Langensiepen, "Trend modelling of elderly lifestyle within an occupancy simulator", in *13th International Conference on Computer Modelling and Simulation (UKSim)*. 2011, pp. 156–161, IEEE. 10

[27] Julie Soulas, Philippe Lenca, and André Thépaut, "Unsupervised discovery of activities of daily living characterized by their periodicity and variability", *Engineering Applications of Artificial Intelligence*, vol. 45, pp. 90–102, 2015. 10

[28] G Virone and A Sixsmith, "Monitoring activity patterns and trends of older adults", in *30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2008, pp. 2071–2074. 10

[29] John Kemp, Elena Gaura, Ramona Rednic, and James Brusey, "Long-term behavioural change detection through pervasive sensing", in *14th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*. IEEE, 2013, pp. 629–634. 10

[30] NK Suryadevara, Subhas C. Mukhopadhyay, R. Wang, and RK Rayudu, "Forecasting the behavior of an elderly using wireless sensors data in a smart home", *Engineering Applications of Artificial Intelligence*, vol. 26, no. 10, pp. 2641–2652, 2013. 11

[31] María Ros, Manuel Pegalajar, Miguel Delgado, Amparo Vila, Derek T Anderson, James M Keller, and Mihail Popescu, "Linguistic summarization of long-term trends for understanding change in human behavior", in *IEEE International Conference on Fuzzy Systems*, 2011, pp. 2080–2087. 11

[32] Abdur Rahim Mohammad Forkan, Ibrahim Khalil, Zahir Tari, Sebti Foufou, and Abdelaziz Bouras, "A context-aware approach for long-term behavioural change detection and abnormality prediction in ambient assisted living", *Pattern Recognition*, vol. 48, no. 3, pp. 628–641, 2015. 11

[33] Robert M Hirsch, Richard B Alexander, and Richard A Smith, "Selection of methods for the detection and estimation of trends in water quality", *Water Resources Research*, vol. 27, no. 5, pp. 803–813, 1991. 11

[34] John Pfister and Chia-Hung Chou, "Progress in reducing mortality among persons 65 to 74 years of age in wisconsin", *WMJ-MADISON-*, vol. 102, no. 8, pp. 52–56, 2003. 12

[35] Céline Chadefaux, Anne-Solenn Le Hô, Ludovic Bellot-Gurlet, and Ina Reiche, "Curve-fitting micro-atr-ftir studies of the amide i and ii bands of type i collagen in archaeological bone materials", *e-PRESERVATION Science*, vol. 6, pp. 129–137, 2009. 12

[36] Robert M Hirsch, James R Slack, and Richard A Smith, "Techniques of trend analysis for monthly water quality data", *Water Resources Research*, vol. 18, no. 1, pp. 107–121, 1982. 12

[37] Sander Greenland and Matthew P Longnecker, "Methods for trend estimation from summarized dose-response data, with applications to meta-analysis", *American Journal of Epidemiology*, vol. 135, no. 11, pp. 1301–1309, 1992. 12

[38] Margaret A Dudeck, Lindsey M Weiner, Katherine Allen-Bridson, Paul J Malpiedi, Kelly D Peterson, Daniel A Pollock, Dawn M Sievert, and Jonathan R Edwards, "National healthcare safety network (nhsn) report, data summary for 2012, device-associated module", *American Journal of Infection Control*, vol. 41, no. 12, pp. 1148, 2013. 12

[39] Janina Mozejko, *Detecting and Estimating Trends of Water Quality Parameters*, INTECH Open Access Publisher, 2012. 12

[40] Rohit Moghe and Mirrasoul J Mousavi, "Trend analysis techniques for incipient fault prediction", in *Power & Energy Society General Meeting, 2009. PES'09. IEEE*, 2009, pp. 1–8. 12

[41] D Rosenberg, "Trend analysis and interpretation", *Key Concepts and Methods for Maternal and Child Health Professionals*, 1997. 13

[42] Haoming Huang, Michel Pasquier, and Chai Quek, "Financial market trading system with a hierarchical coevolutionary fuzzy predictive model", *IEEE Transactions on Evolutionary Computation*, vol. 13, no. 1, pp. 56–70, 2009. 13

[43] Martin Štěpnička, Paulo Cortez, Juan Peralta Donate, and Lenka Štěpničková, "Forecasting seasonal time series with computational intelligence: on recent methods and the potential of their combinations", *Expert Systems with Applications*, vol. 40, no. 6, pp. 1981–1992, 2013. 13

[44] G Rubell Marion Lincy and C Jessy John, "A multiple fuzzy inference systems framework for daily stock trading with application to nasdaq stock exchange", *International Journal of Expert Systems with Applications*, vol. 44, no. C, pp. 13–21, 2016. 13

[45] Konstandinos Chourmouziadis and Prodromos D Chatzoglou, "An intelligent short term stock trading fuzzy system for assisting investors in portfolio management", *Expert Systems with Applications*, vol. 43, pp. 298–311, 2016. 13

[46] Nedaa Agami, Amir Atiya, Mohamed Saleh, and Hisham El-Shishiny, "A neural network based dynamic forecasting model for trend impact analysis", *Technological Forecasting and Social Change*, vol. 76, no. 7, pp. 952–962, 2009. 13

[47] George S Atsalakis, Eftychios E Protopapadakis, and Kimon P Valavanis, "Stock trend forecasting in turbulent market periods using neuro-fuzzy systems", *Operational Research*, pp. 1–25, 2015. 13

[48] Labiba Gillani Fahad, Arshad Ali, and Muttukrishnan Rajarajan, "Long term analysis of daily activities in smart home.", in *Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning.*, 2013. 13

[49] Bruno Lepri, Albert Ali Salah, Fabio Pianesi, and Alex Sandy Pentland, *Human behavior understanding for inducing behavioral change: Social and theoretical aspects*, pp. 252–263, Constructing Ambient Intelligence. Springer, 2012. 14

[50] JC SanMiguel, Marcos Escudero-Viñolo, José M Martínez, and Jesús Bescós, "Real-time single-view video event recognition in controlled environments", in *9th International Workshop on Content-Based Multimedia Indexing*. IEEE, 2011, pp. 91–96. 14

[51] Jake K Aggarwal and Lu Xia, "Human activity recognition from 3d data: A review", *Pattern Recognition Letters*, vol. 48, pp. 70–80, 2014. 14

[52] Diane J Cook, Narayanan C Krishnan, and Parisa Rashidi, "Activity discovery and activity recognition: A new partnership", *IEEE Transactions on Cybernetics*, vol. 43, no. 3, pp. 820–828, 2013. 14

[53] Asma Benmansour, Abdelhamid Bouchachia, and Mohamed Feham, "Human activity recognition in pervasive single resident smart homes: State of art", in *12th International Symposium on Programming and Systems*. IEEE, 2015, pp. 1–9. 14

[54] JP Black, W. Segmuller, N. Cohen, B. Leiba, A. Misra, MR Ebling, and E. Stern, "Pervasive computing in health care: Smart spaces and enterprise information systems", in *Proceedings of the MobiSys Workshop on Context Awareness*, 2004. 14

[55] Fariba Sadri, "Ambient intelligence: A survey", *ACM Computing Surveys (CSUR)*, vol. 43, no. 4, pp. 36, 2011. 14

[56] Varun Chandola, Arindam Banerjee, and Vipin Kumar, "Anomaly detection for discrete sequences: A survey", *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 5, pp. 823–839, 2012. 14

[57] Damla Arifoglu and Abdelhamid Bouchachia, "Activity recognition and abnormal behaviour detection with recurrent neural networks", *Procedia Computer Science*, vol. 110, pp. 86–93, 2017. 14

[58] Xiaohui Liang, Mrinmoy Barua, Le Chen, Rongxing Lu, Xuemin Shen, Xu Li, and Henry Y Luo, "Enabling pervasive healthcare through continuous remote health monitoring", *IEEE Wireless Communications*, vol. 19, no. 6, 2012. 14

[59] CD Nugent, MD Mulvenna, Xin Hong, and S Devlin, "Experiences in the development of a smart lab", *International Journal of Biomedical Engineering and Technology*, vol. 2, no. 4, pp. 319–331, 2009. 14

[60] Liming Chen and Ismail Khalil, *Activity recognition: Approaches, practices and trends*, pp. 1–31, Activity Recognition in Pervasive Intelligent Environments. Springer, 2011. 14

[61] Sawsan Mousa Mahmoud, *Identification and Prediction of Abnormal Behaviour Activities of Daily Living in Intelligent Environments*, PhD thesis, Nottingham Trent University, 2012. 15

[62] Giovanni Acampora, Diane J Cook, Parisa Rashidi, and Athanasios V Vasilakos, "A survey on ambient intelligence in healthcare", *Proceedings of the IEEE*, vol. 101, no. 12, pp. 2470–2494, 2013. 15

[63] Zhen Li, Zhiqiang Wei, Yaofeng Yue, Hao Wang, Wenyan Jia, Lora E Burke, Thomas Baranowski, and Mingui Sun, "An adaptive hidden markov model for activity recognition based on a wearable multi-sensor device", *Journal of Medical Systems*, vol. 39, no. 5, pp. 57, 2015. 15

[64] Andreas Bulling, Ulf Blanke, and Bernt Schiele, "A tutorial on human activity recognition using body-worn inertial sensors", *ACM Computing Surveys (CSUR)*, vol. 46, no. 3, pp. 33, 2014. 15

[65] Narayanan C Krishnan and Diane J Cook, "Activity recognition on streaming sensor data", *Pervasive and Mobile Computing*, vol. 10, pp. 138–154, 2014. 15

[66] Stefan Oniga and Jozsef Suto, "Human activity recognition using neural networks", in *15th International Carpathian Control Conference*. IEEE, 2014, pp. 403–406. 15

[67] Diane J Cook, "Learning setting-generalized activity models for smart spaces", *IEEE Intelligent Systems*, vol. 27, no. 1, pp. 32–38, 2012. 15

[68] Jose Antonio Iglesias, Plamen Angelov, Agapito Ledezma, and Araceli Sanchis, "Human activity recognition based on evolving fuzzy systems", *International Journal of Neural Systems*, vol. 20, no. 05, pp. 355–364, 2010. 15

[69] Juan Ye, Simon Dobson, and Susan McKeever, "Situation identification techniques in pervasive computing: A review", *Pervasive and Mobile Computing*, vol. 8, no. 1, pp. 36–66, 2012. 16

[70] T. V. Duong, H. H. Bui, D. Q. Phung, and S. Venkatesh, "Activity recognition and abnormality detection with the switching hidden semi-markov model", in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005, vol. 1, pp. 838–845 vol. 1, ID: 1. 16

[71] Dorra Trabelsi, Samer Mohammed, Faicel Chamroukhi, Latifa Oukhellou, and Yacine Amirat, "An unsupervised approach for automatic activity recognition based on hidden markov model regression", *IEEE Transactions*

on Automation Science and Engineering, vol. 10, no. 3, pp. 829–835, 2013. 16

[72] Eunju Kim, S. Helal, and D. Cook, "Human activity recognition and pattern discovery", *Pervasive Computing, IEEE*, vol. 9, no. 1, pp. 48–53, 2010, ID: 1. 16

[73] Katherine Brittain, Lynne Corner, Louise Robinson, and John Bond, "Ageing in place and technologies of place: the lived experience of people with dementia in changing social, physical and technological environments", *Sociology of Health & Illness*, vol. 32, no. 2, pp. 272–287, 2010. 16

[74] Thomas Dyhre Nielsen and Finn Verner Jensen, *Bayesian networks and decision graphs*, Springer, 2009. 16

[75] Youngwook Kim and Hao Ling, "Human activity classification based on micro-doppler signatures using a support vector machine", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 5, pp. 1328–1337, 2009. 16

[76] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, and Jorge L Reyes-Ortiz, "Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine", in *International Workshop on Ambient Assisted Living*. Springer, 2012, pp. 216–223. 16

[77] Sawsan Mahmoud, Ahmad Lotfi, and Caroline Langensiepen, "Behavioural pattern identification and prediction in intelligent environments", *Applied Soft Computing*, vol. 13, no. 4, pp. 1813–1822, 2013. 16

[78] Jianbo Yang, Minh Nhut Nguyen, Phyo Phyo San, Xiaoli Li, and Shonali Krishnaswamy, "Deep convolutional neural networks on multichannel time series for human activity recognition.", in *the Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015, pp. 3995–4001. 16

[79] Natalia Díaz Rodríguez, Manuel P Cuéllar, Johan Lilius, and Miguel Delgado Calvo-Flores, "A fuzzy ontology for semantic modelling and recogni-

tion of human behaviour", *Knowledge-Based Systems*, vol. 66, pp. 46–60, 2014. 16

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

[81] Mohammad Helmi and Seyed Mohammad T AlModarresi, "Human activity recognition using a fuzzy inference system", in *IEEE International Conference on Fuzzy Systems*, 2009, pp. 1897–1902. 16

[82] Bo Yao, Hani Hagras, Mohammed J Alhaddad, and Daniyal Alghazzawi, "A fuzzy logic-based system for the automation of human behavior recognition using machine vision in intelligent environments", *Soft Computing*, vol. 19, no. 2, pp. 499–506, 2015. 16

[83] Laurens Cherchye, Timo Kuosmanen, et al., "Benchmarking sustainable development: A synthetic meta-index approach", *Documento preparado en el contexto del programa de investigación en Métodos No paramétricos en Enomomía de la Producción, los recursos Naturales y el Medio Ambiente*, 2002. 16

[84] Earl R Babbie, *The practice of social research*, Wadsworth Publishing Company, 1989. 17

[85] Michael Hopkins, "Human development revisited: A new undp report", *World Development*, vol. 19, no. 10, pp. 1469–1473, 1991. 17

[86] Laurens Cherchye, CA Knox Lovell, Wim Moesen, and Tom Van Puyenbroeck, "One market, one number? a composite indicator assessment of eu internal market dynamics", *European Economic Review*, vol. 51, no. 3, pp. 749–779, 2007. 18

[87] Giovanni Cerulli and Andrea Filippetti, "The complementary nature of technological capabilities: Measurement and robustness issues", *Techno-*

logical Forecasting and Social Change, vol. 79, no. 5, pp. 875–887, 2012. 18

[88] William Leigh, Russell Purvis, and James M Ragusa, "Forecasting the nyse composite index with technical analysis, pattern recognizer, neural network, and genetic algorithm: a case study in romantic decision support", *Decision Support Systems*, vol. 32, no. 4, pp. 361–377, 2002. 18

[89] Yakup Kara, Melek Acar Boyacioglu, and Ömer Kaan Baykan, "Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the istanbul stock exchange", *Expert Systems with Applications*, vol. 38, no. 5, pp. 5311–5319, 2011. 18

[90] Cagdas Hakan Aladag, Ufuk Yolcu, Erol Egrioglu, and Ali Z Dalar, "A new time invariant fuzzy time series forecasting method based on particle swarm optimization", *Applied Soft Computing*, vol. 12, no. 10, pp. 3291–3299, 2012. 18

[91] Enjian Bai, Wai Keung Wong, WC Chu, Min Xia, and Feng Pan, "A heuristic time-invariant model for fuzzy time series forecasting", *Expert Systems with Applications*, vol. 38, no. 3, pp. 2701–2707, 2011. 18

[92] Shou-Hsiung Cheng, Shyi-Ming Chen, and Wen-Shan Jian, "Fuzzy time series forecasting based on fuzzy logical relationships and similarity measures", *Information Sciences*, vol. 327, pp. 272–287, 2016. 18

[93] Oscar D Lara and Miguel A Labrador, "A survey on human activity recognition using wearable sensors.", *IEEE Communications Surveys and Tutorials*, vol. 15, no. 3, pp. 1192–1209, 2013. 19

[94] Manish Gupta, Jing Gao, Charu C Aggarwal, and Jiawei Han, "Outlier detection for temporal data: A survey", *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 9, pp. 2250–2267, 2014. 19

[95] Taketoshi Mori, Akinori Fujii, Masamichi Shimosaka, Hiroshi Noguchi, and Tomomasa Sato, "Typical behavior patterns extraction and anomaly detection algorithm based on accumulated home sensor data", in *Future gen-*

eration communication and networking (FGCN 2007). IEEE, 2007, vol. 2, pp. 12–18. 19

[96] Jaewan Shin, Dongkyoo Shin, and DongIl Shin, "Predicting of abnormal behavior using hierarchical markov model based on user profile in ubiquitous environment", in *International Conference on Grid and Pervasive Computing*. Springer, 2013, pp. 612–622. 19

[97] Svebor Karaman, Jenny Benois-Pineau, Vladislavs Dovgalecs, Rémi Mégret, Julien Pinquier, Régine André-Obrecht, Yann Gaëstel, and Jean-François Dartigues, "Hierarchical hidden markov model in detecting activities of daily living in wearable videos for studies of dementia", *Multimedia Tools and Applications*, vol. 69, no. 3, pp. 743–771, 2014. 19

[98] Panagiota Antonakaki, Dimitrios Kosmopoulos, and Stavros J Perantonis, "Detecting abnormal human behaviour using multiple cameras", *Signal Processing*, vol. 89, no. 9, pp. 1723–1738, 2009. 20

[99] Matthew Field, David Stirling, Zengxi Pan, Montserrat Ros, and Fazel Naghdy, "Recognizing human motions through mixture modeling of inertial data", *Pattern Recognition*, vol. 48, no. 8, pp. 2394–2406, 2015. 20

[100] Barbara Bruno, Fulvio Mastrogiovanni, Antonio Sgorbissa, Tullio Vernazza, and Renato Zaccaria, "Analysis of human behavior recognition algorithms based on acceleration data", in *IEEE International Conference on Robotics and Automation*, 2013, pp. 1602–1607. 20

[101] Barbara Bruno, Fulvio Mastrogiovanni, Antonio Sgorbissa, Tullio Vernazza, and Renato Zaccaria, "Human motion modelling and recognition: A computational approach", in *IEEE International Conference on Automation Science and Engineering*, 2012, pp. 156–161. 20

[102] Fabien Cardinaux, Simon Brownsell, Mark Hawley, and David Bradley, "Modelling of behavioural patterns for abnormality detection in the context of lifestyle reassurance", in *Iberoamerican Congress on Pattern Recognition*. Springer, 2008, pp. 243–251. 20

[103] Vikramaditya Jakkula and Diane Cook, "Outlier detection in smart environment structured power datasets", in *Sixth International Conference on Intelligent Environments*, 2010, pp. 29–33. 20

[104] Vikramaditya Jakkula, Diane J Cook, and Aaron S Crandall, "Temporal pattern discovery for anomaly detection in a smart home", in *International Conference on Intelligent Environments*. 2007, pp. 339–345, IET. 20

[105] AK Bourke, JV Obrien, and GM Lyons, "Evaluation of a threshold-based tri-axial accelerometer fall detection algorithm", *Gait & Posture*, vol. 26, no. 2, pp. 194–199, 2007. 21

[106] P Jantaraprim, P Phukpattaranont, C Limsakul, and B Wongkittisuksa, "Improving the accuracy of a fall detection algorithm using free fall characteristics", in *International Conference on Electrical Engineering/Electronics Computer Telecommunications and Information Technology*. IEEE, 2010, pp. 501–504. 21

[107] Homay Danaei Mehr, Huseyin Polat, and Aydin Cetin, "Resident activity recognition in smart homes by using artificial neural networks", in *4th International Istanbul Smart Grid Congress and Fair*. IEEE, 2016, pp. 1–5. 21

[108] Sebastian Lühr, Geoff West, and Svetha Venkatesh, "Recognition of emergent human behaviour in a smart home: A data mining approach", *Pervasive and Mobile Computing*, vol. 3, no. 2, pp. 95–116, 2007. 21

[109] Simona D'Oca and Tianzhen Hong, "A data-mining approach to discover patterns of window opening and closing behavior in offices", *Building and Environment*, vol. 82, pp. 726–739, 2014. 22

[110] Maimaitimin Maierdan, Keigo Watanabe, and Shoichi Maeyama, "Human behavior recognition system based on 3-dimensional clustering methods", in *13th International Conference on Control, Automation and Systems*. IEEE, 2013, pp. 1133–1137. 22

[111] Qing Zhang, Mohan Karunanithi, Dana Bradford, and Yasmin van Kasteren, "Activity of daily living assessment through wireless sensor data", in *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014, pp. 1752–1755. 22

[112] Yan Yan, Elisa Ricci, Gaowen Liu, and Nicu Sebe, "Egocentric daily activity recognition via multitask clustering", *IEEE Transactions on Image Processing*, vol. 24, no. 10, pp. 2984–2995, 2015. 22

[113] Li Liu, Shu Wang, Yuxin Peng, Zigang Huang, Ming Liu, and Bin Hu, "Mining intricate temporal rules for recognizing complex activities of daily living under uncertainty", *Pattern Recognition*, vol. 60, pp. 1015–1028, 2016. 22

[114] Rui Hu, Hieu Pham, Philipp Buluschek, and Daniel Gatica-Perez, "Elderly people living alone: Detecting home visits with ambient and wearable sensing", in *In Proceedings of MMHealth*, 2017. 23

[115] Miao Yu, Adel Rhuma, Syed Mohsen Naqvi, Liang Wang, and Jonathon Chambers, "A posture recognition-based fall detection system for monitoring an elderly person in a smart home environment", *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 6, pp. 1274–1286, 2012. 23

[116] Vikramaditya R Jakkula and Diane J Cook, "Detecting anomalous sensor events in smart home data for enhancing the living experience.", *Artificial Intelligence and Smarter Living*, vol. 11, no. 201, pp. 1, 2011. 23

[117] Bo Yao, Hani Hagras, Mohammed J Alhaddad, and Daniyal Alghazzawi, "A fuzzy logic-based system for the automation of human behavior recognition using machine vision in intelligent environments", *Soft Computing*, vol. 19, no. 2, pp. 499–506, 2015. 23

[118] Moshe Blank, Lena Gorelick, Eli Shechtman, Michal Irani, and Ronen Basri, "Actions as space-time shapes", in *Tenth IEEE International Conference on Computer Vision*, 2005, vol. 2, pp. 1395–1402. 24

[119] JL Castro, Miguel Delgado, Javier Medina, and MD Ruiz-Lozano, "Intelligent surveillance system with integration of heterogeneous information for intrusion detection", *Expert Systems with Applications*, vol. 38, no. 9, pp. 11182–11192, 2011. 24

[120] Giovanni Acampora, Pasquale Foggia, Alessia Saggese, and Mario Vento, "A hierarchical neuro-fuzzy architecture for human behavior analysis", *Information Sciences*, vol. 310, pp. 130–148, 2015. 24

[121] Halldór Janetzko, Florian Stoffel, Sebastian Mittelstädt, and Daniel A Keim, "Anomaly detection for visual analytics of power consumption data", *Computers & Graphics*, vol. 38, pp. 27–37, 2014. 24, 25

[122] Juan Luis Carús Candás, Víctor Peláez, Gloria López, Miguel Ángel Fernández, Eduardo Álvarez, and Gabriel Díaz, "An automatic data mining method to detect abnormal human behaviour using physical activity measurements", *Pervasive and Mobile Computing*, vol. 15, pp. 228–241, 2014. 25

[123] Kyungseo Park, Yong Lin, Vangelis Metsis, Zhengyi Le, and Fillia Makedon, "Abnormal human behavioral pattern detection in assisted living environments", in *Proceedings of the 3rd International Conference on PErvasive Technologies Related to Assistive Environments*. ACM, 2010, p. 9. 25

[124] Lei Meng, Chunyan Miao, and Cyril Leung, "Towards online and personalized daily activity recognition, habit modeling, and anomaly detection for the solitary elderly through unobtrusive sensing", *Multimedia Tools and Applications*, vol. 76, no. 8, pp. 10779–10799, 2017. 25

[125] RP Jagadeesh Chandra Bose, Wil MP Van Der Aalst, Indre Zliobaite, and Mykola Pechenizkiy, "Dealing with concept drifts in process mining", *IEEE transactions on neural networks and learning systems*, vol. 25, no. 1, pp. 154–171, 2014. 26

[126] Indrė Žliobaitė, Mykola Pechenizkiy, and Joao Gama, "An overview of concept drift applications", in *Big Data Analysis: New Algorithms for a New Society*, pp. 91–114. Springer, 2016. 26

[127] João Gama, Indrė Žliobaitė, Albert Bifet, Mykola Pechenizkiy, and Abdelhamid Bouchachia, "A survey on concept drift adaptation", *ACM computing surveys (CSUR)*, vol. 46, no. 4, pp. 44, 2014. 26

[128] András Horányi, Carla Cardinali, and Luca Centurioni, "The global numerical weather prediction impact of mean-sea-level pressure observations from drifting buoys", *Quarterly Journal of the Royal Meteorological Society*, vol. 143, no. 703, pp. 974–985, 2017. 26

[129] A Al-Habaibeh, Fangmin Shi, Neil Brown, David Kerr, M Jackson, and Robert M Parkin, "A novel approach for quality control system using sensor fusion of infrared and visual image processing for laser sealing of food containers", *Measurement Science and Technology*, vol. 15, no. 10, pp. 1995, 2004. 26

[130] F Edthofer, J Van den Broeke, J Ettl, W Lettl, and A Weingartner, "Reliable online water quality monitoring as basis for fault tolerant control", in *Control and Fault-Tolerant Systems (SysTol), 2010 Conference on*. IEEE, 2010, pp. 57–62. 26

[131] Jessamyn Dahmen, Brian L Thomas, Diane J Cook, and Xiaobo Wang, "Activity learning as a foundation for security monitoring in smart homes", *Sensors*, vol. 17, no. 4, pp. 737, 2017. 30

[132] Carlos Ramos, Juan Carlos Augusto, and Daniel Shapiro, "Ambient intelligencethe next step for artificial intelligence", *IEEE Intelligent Systems*, vol. 23, no. 2, pp. 15–18, 2008. 31

[133] Emile Aarts and Reiner Wichert, "Ambient intelligence", in *Technology Guide*, pp. 244–249. Springer, 2009. 32

[134] JustChecking monitoring system, "Supporting independence people with dementia", 2015. 33

[135] Mohammad Javad Akhlaghinia, *Occupancy monitoring and prediction in ambient intelligent environment*, PhD thesis, Nottingham Trent University, 2010. 34

[136] Saifullizam Puteh, *User profiling in the intelligent office*, PhD thesis, Nottingham Trent University, 2013. 34

[137] SB Kotsiantis, D Kanellopoulos, and PE Pintelas, "Data preprocessing for supervised leaning", *International Journal of Computer Science*, vol. 1, no. 2, pp. 111–117, 2006. 35

[138] Richard Chandler and Marian Scott, *Statistical methods for trend detection and analysis in the environmental sciences*, John Wiley & Sons, 2011. 35, 60

[139] Julian PT Higgins, Sally Green, et al., *Cochrane handbook for systematic reviews of interventions*, vol. 5, Wiley Online Library, 2008. 38

[140] James L Peugh and Craig K Enders, "Missing data in educational research: A review of reporting practices and suggestions for improvement", *Review of Educational Research*, vol. 74, no. 4, pp. 525–556, 2004. 39

[141] Paul D Allison, *Missing data*, vol. 136, Sage publications, 2001. 39

[142] Arpad Gellert and Lucian Vintan, "Person movement prediction using hidden markov models", *Studies in Informatics and Control*, vol. 15, no. 1, pp. 17, 2006. 42

[143] Norbert Noury and Tareq Hadidi, "Computer simulation of the activity of the elderly person living independently in a health smart home", *Computer Methods and Programs in Biomedicine*, vol. 108, no. 3, pp. 1216–1228, 2012. 43

[144] Balázs Kormányos and Béla Pataki, "Multilevel simulation of daily activities: Why and how?", in *Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA), 2013 IEEE International Conference on*. IEEE, 2013, pp. 1–6. 44

[145] Fabien Cardinaux, Simon Brownsell, David Bradley, and Mark S Hawley, "A home daily activity simulation model for the evaluation of lifestyle monitoring systems", *Computers in biology and medicine*, vol. 43, no. 10, pp. 1428–1436, 2013. 44

[146] Francis J Alexander and Alejandro L Garcia, "The direct simulation monte carlo method", *Computers in Physics*, vol. 11, no. 6, pp. 588, 1997. 47

[147] ES Oran, CK Oh, and BZ Cybyk, "Direct simulation monte carlo: Recent advances and applications 1", *Annual Review of Fluid Mechanics*, vol. 30, no. 1, pp. 403–441, 1998. 47

[148] JPC Kleijen, "Validation of models: statistical techniques and data availability", in *Simulation Conference Proceedings*. IEEE, 1999, vol. 1, pp. 647–654. 50

[149] Jerry L Hintze and Ray D Nelson, "Violin plots: a box plot-density trace synergism", *The American Statistician*, vol. 52, no. 2, pp. 181–184, 1998. 55

[150] Guy P Nason, "Stationary and non-stationary time series", *Statistics in Volcanology*, 2006. 60

[151] Zhaohua Wu, Norden E Huang, Steven R Long, and Chung-Kang Peng, "On the trend, detrending, and variability of nonlinear and nonstationary time series", *Proceedings of the National Academy of Sciences*, vol. 104, no. 38, pp. 14889–14894, 2007. 60

[152] Richard O. Gilbert, *Statistical methods for environmental pollution monitoring*, John Wiley & Sons, 1987. 60

[153] Charles C Holt, "Forecasting seasonals and trends by exponentially weighted moving averages", *International Journal of Forecasting*, vol. 20, no. 1, pp. 5–10, 2004. 61, 64

[154] SW Roberts, "Control chart tests based on geometric moving averages", *Technometrics*, vol. 1, no. 3, pp. 239–250, 1959. 61

[155] Clif Droke, *Moving averages simplified*, Marketplace Books, 2001. 61, 62

[156] Engineering Statistics Handbook, "Single exponential smoothing", 2013. 62, 65

[157] Everette S Gardner, "Exponential smoothing: The state of the artpart ii", *International Journal of Forecasting*, vol. 22, no. 4, pp. 637–666, 2006. 64

[158] David R Cox, "Prediction by exponentially weighted moving averages and related methods", *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 414–422, 1961. 66

[159] Perry J Kaufman, *Commodity trading systems and methods*, Wiley, 1978. 67

[160] Gonzalo R Arce, *Nonlinear signal processing: a statistical approach*, John Wiley & Sons, 2005. 68

[161] Robert M Hirsch and James R Slack, "A nonparametric trend test for seasonal data with serial dependence", *Water Resources Research*, vol. 20, no. 6, pp. 727–732, 1984. 69, 71

[162] Graham McBride, "Anomalies and remedies in non-parametric seasonal trend tests and estimates", *National Institute of Water and Atmospheric Research, Hamilton*, 2000. 71

[163] Gerald Belle and James P Hughes, "Nonparametric tests for trend in water quality", *Water Resources Research*, vol. 20, no. 1, pp. 127–136, 1984. 71

[164] F James Rohlf and Robert R Sokal, *Statistical tables*, Macmillan, 1995. 71

[165] J-SR Jang, "Anfis: adaptive-network-based fuzzy inference system", *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 23, no. 3, pp. 665–685, 1993. 73

[166] Melek Acar Boyacioglu and Derya Avci, "An adaptive network-based fuzzy inference system (anfis) for the prediction of stock market return: the case of the istanbul stock exchange", *Expert Systems with Applications*, vol. 37, no. 12, pp. 7908–7912, 2010. 73

[167] Gerald Appel, *Technical analysis: power tools for active investors*, FT Press, 2005. 77

[168] Wing-Keung Wong, Meher Manzur, and Boon-Kiat Chew, "How rewarding is technical analysis evidence from singapore stock market", *Applied Financial Economics*, vol. 13, no. 7, pp. 543–551, 2003. 77, 78

[169] R Rosillo, D De la Fuente, and JAL Brugos, "Technical analysis and the spanish stock exchange: testing the rsi, macd, momentum and stochastic rules using spanish market companies", *Applied Economics*, vol. 45, no. 12, pp. 1541–1550, 2013. 77

[170] Nguyen Hoang Hung, "Various moving average convergence divergence trading strategies: a comparison", *Investment Management and Financial Innovations*, vol. 13, no. 2, 2016. 77

[171] Harry Boxer, "Moving average convergence/divergence", *Profitable Day and Swing Trading: Using Price/Volume Surges and Pattern Recognition to Catch Big Moves in the Stock Market*, pp. 91–102, 2014. 78

[172] John J Murphy, *The visual investor: how to spot market trends*, vol. 443, John Wiley & Sons, 2009. 79

[173] Terence Tai-Leung Chong and Wing-Kam Ng, "Technical analysis and the london stock exchange: testing the "macd" and "rsi" rules using the "ft30"", *Applied Economics Letters*, vol. 15, no. 14, pp. 1111–1114, 2008. 80

[174] J Wells Wilder Jr, "The relative strength index ", *Journal of Technical Analysis of Stocks and Commodities*, vol. 4, pp. 343–346, 1986. 80

[175] Sawsan Mahmoud, Ahmad Lotfi, and Caroline Langensiepen, "User activities outliers detection; integration of statistical and computational intelligence techniques", *Computational Intelligence*, 2014. 110

[176] Andong Zhan, Max A Little, Denzil A Harris, Solomon O Abiola, E Dorsey, Suchi Saria, and Andreas Terzis, "High frequency remote monitoring of parkinson's disease via smartphone: Platform overview and medication response detection", *Computers and Society*, 2016. 110

[177] B. Wallace, F. Knoefel, R. Goubran, P. Masson, A. Baker, B. Allard, V. Guana, and E. Stroulia, "Detecting cognitive ability changes in patients with moderate dementia using a modified 'whack-a-mole' game", *IEEE Transactions on Instrumentation and Measurement*, vol. PP, no. 99, pp. 1–14, 2017. 110

[178] W. D. Kearns, J. L. Fozard, and V. O. Nams, "Movement path tortuosity in free ambulation: Relationships to age and brain disease", *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 2, pp. 539–548, March 2017. 110

[179] Ehsan Nazerfard and Diane J Cook, "Crafft: an activity prediction model based on bayesian networks", *Journal of Ambient Intelligence and Humanized Computing*, vol. 6, no. 2, pp. 193–205, 2015. 110

[180] Pau-Choo Chung and Chin-De Liu, "A daily behavior enabled hidden markov model for human behavior understanding", *Pattern Recognition*, vol. 41, no. 5, pp. 1572–1580, 2008. 110

[181] Eying Wu, Peng Zhang, Tun Lu, Hansu Gu, and Ning Gu, "Behavior prediction using an improved hidden markov model to support people with disabilities in smart homes", in *20th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*. IEEE, 2016, pp. 560–565. 110

[182] Fco Javier Ordóñez, Paula de Toledo, and Araceli Sanchis, "Sensor-based bayesian detection of anomalous living patterns in a home setting", *Personal and Ubiquitous Computing*, vol. 19, no. 2, pp. 259–270, 2015. 110

[183] Tim van Kasteren and Ben Krose, "Bayesian activity recognition in residence for elders", in *3rd IET International Conference on Intelligent Environments*. 2007, IET. 110

[184] Jie Yin, Qiang Yang, and J. J. Pan, "Sensor-based abnormal human-activity detection", *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 8, pp. 1082–1090, 2008, ID: 1. 111

[185] Zhenyu He and Lianwen Jin, “Activity recognition from acceleration data based on discrete consine transform and svm”, in *International Conference on Systems, Man and Cybernetics*. IEEE, 2009, pp. 5041–5044. 111

[186] Diane J. Cook, Juan C. Augusto, and Vikramaditya R. Jakkula, “Ambient intelligence: Technologies, applications, and opportunities”, *Pervasive and Mobile Computing*, vol. 5, no. 4, pp. 277–298, 2009. 111

[187] Xavier Rafael-Palou, Eloisa Vargiu, and Felip Miralles, “Monitoring people that need assistance through a sensor-based system: Evaluation and first results”, in *Proceedings of Artificial Intelligence and Assistive Medicine*, 2015, vol. 1389. 111

[188] Ahmad Lotfi, Caroline Langensiepen, Pedro A Moreno, Enrique J Gómez, and Saisakul Chernbumroong, “An ambient assisted living technology platform for informal carers of the elderly-icarer”, *EAI Endorsed Transactions on Pervasive Health and Technology*, vol. 3, no. 9, pp. 152393, 2017. 111

[189] Abdur Forkan, Ibrahim Khalil, Ayman Ibaida, and Zahir Tari, “Bdcam: Big data for context-aware monitoring-a personalized knowledge discovery framework for assisted healthcare”, *IEEE Transactions on Cloud Computing*, 2015. 111

[190] Jeremie Saives, Clement Pianon, and Gregory Faraut, “Activity discovery and detection of behavioral deviations of an inhabitant from binary sensors”, *IEEE Transactions on Automation Science and Engineering*, vol. 12, no. 4, pp. 1211–1224, 2015. 111

[191] OECD, “Statistics, knowledge and policy 2007”, `http://dx.doi.org/10.1787/9789264043244-en`, 2008. 111

[192] Peng Zhou, BW Ang, and DQ Zhou, “Weighting and aggregation in composite indicator construction: A multiplicative optimization approach”, *Social Indicators Research*, vol. 96, no. 1, pp. 169–181, 2010. 112