May 2020

Identification of Patterns and Disruptions in Ambient Sensor Data from Private Homes

Yan Wang
University of South Florida

Identification of Patterns and Disruptions in Ambient Sensor Data from Private Homes

by

Yan Wang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Industrial and Management Systems Engineering
College of Engineering
University of South Florida

Major Professor: Ali Yalcin, Ph.D.
Carla VandeWeerd, Ph.D.
Mingyang Li, Ph.D.
Yu Zhang, Ph.D.
Lu Lu, Ph.D.

Date of Approval:
March 6, 2020

Keywords: Age in Place, Health Monitoring, Mobility, Predictability, Smart Homes

Dedication

I dedicate my dissertation work to my parents.

Acknowledgments

I would like to express my appreciation to my mentor Dr. Ali Yalcin, Associate Professor, Department of Industrial and Management Systems Engineering, for his patient guidance and assistance through each stage of the process, and the valuable time, energy and endeavor he spent on cultivating my skills and critical thinking. Without his constant support and encouragement throughout the years, this research would not have been completed. His intelligence, sincerity, enthusiasm for research, hardworking spirit, and kindness have set a role model to me and inspired me to greater efforts.

I would express my appreciation to Dr. Carla VandeWeerd, Dr. José Zayas-Castro, and Dr. Tapas K. Das. Their sincere words, advice, and unconditional support motivate me to move ahead bravely. I would also thank the committee, Dr. Mingyang Li, Dr. Yu Zhang, and Dr. LuLu, for their valuable time, guidance, and help over the past years.

I would also like to extend my gratitude to the exceptional team members in HomeSense for their selfless sharing and help in this research. Thanks also go out to the faculty of the Department of Industrial and Management Systems Engineering. Working with them is a valuable experience for me in my graduate career.

Lastly, I would like to thank all my friends for the company.

Table of Contents

List of Tables

List of Figures

Abstract

The world's population is rapidly aging and the increasing demand for home and health care services from this aging population brings unprecedented challenges to the economy and society. Ambient-assisted smart homes, residences equipped with ambient sensors to monitor the resident's daily activities in a continuous and unobtrusive way, present great potential to manage the growing care service needs of this older population segment, and enable them to age-in-place.

Despite growing research, using ambient sensor data from private homes to monitor daily activities, health and wellness still faces significant challenges. To study ambient sensor data from private homes where annotated data is unavailable and sensor layouts are variable, we proposed a novel two-phase location and status estimation algorithm to monitor health and wellness related metrics from ambient sensor data. The proposed algorithm is highly accurate as validated by a mobile app that prompts participants with questions about the estimated time of their daily activities. The outputs of this algorithm facilitate the visualization and examination of older adults' daily patterns and activities, and through case studies, we show that it has the potential to be used with a wide range of ambient sensor networks with any mix of motion sensor types.

We also studied human mobility in private homes. Understanding human mobility is fundamental and critical for the design of context-aware assistive services in smart homes. We represent the resident's movement trajectory based on ambient motion sensor data and use the entropy rate to quantify the regularity of the resident's mobility patterns to estimate an upper bound of predictability. A change point detection algorithm based on penalized contrast function is used

to identify the time periods when the data does not completely reflect the resident's activities due to the presence of visitors and sensors system faults. Experimental results using data collected from 10 private homes over periods of 178 to 713 days show that human mobility at home is not completely random but regular and highly predictable independent of variations in floor plans and individual daily routines, which is consistent with the conclusions about human mobility in outdoor environments.

Finally, we summarize and analyze records in maintenance logs and bi-weekly assessments about changes and disruptions in ambient sensor data collected from private homes, and suggest potential research directions for the design of stable and reliable health and wellness monitoring systems using ambient sensor systems.

Chapter 1: Introduction

1.1 Note to Reader

Portions of Chapter 1 have been previously accepted by *Health and Technology* and have been reproduced with permission from Springer.

1.2 Background

The growth in the number and proportion of older adults is unprecedented in the history of the United States [1] and the world [2]. According to the United Nations [3], the population of adults over the age of 60 has quadrupled from 205 million in 1950 to almost 810 million in 2012 worldwide. This segment of the population is expected to double in size again - reaching 2 billion persons (more than 20% of the world's population) by 2050. In the US, nearly 1 in every 5 Americans will be an older adult in 2030; and by 2050, more than 89 million Americans will be age 65 and older – double the number in 2010 [1].

As people live longer, the prevalence of chronic conditions is also on the rise. At present, 70 million older adults are suffering from one chronic condition, and 2/3 of adults over the age of 65 are suffering from 2 or more [4]. Chronic diseases result in negative health consequences and "people living with one or more chronic diseases often experience a diminished quality of life, generally reflected by a long period of decline and disability associated with their illness" [5]. The nation's expenditures for health care are already among the highest in developed countries, and, the costs are expected to increase by 55% over the next 10 years as chronic diseases affect the growing numbers of older adults [6]. Today, more than 2/3 of health care costs expended go to treating chronic illnesses, and in older adults, chronic disease treatment accounts for 95% of health care expenditures [7].

The rises in projected health care costs are unsustainable and call for improving the ways in which we manage health [8]. Additionally, institutional health care systems are not prepared to meet the needs of the growing number of seniors, who have expressed a strong desire to “age in place” in their communities [9]. CDC [10] has defined aging in place as “The ability to live in one’s own home and community safely, independently, and comfortably, regardless of age, income or ability level”. Aging in place offers significant benefits including a reduction in health care costs through avoidance of institutionalization [11], improving quality of life [12], increasing independence [13], expanding/maintaining social networks [14], and reducing risks for cognitive decline and adverse mental health [15].

Recently, ambient-assisted smart homes — residences equipped with ambient sensors and computing technology that monitor the activities and well-being of occupants in their homes — are increasingly seen as facilitating innovative and supportive environments for enabling the healthy, safe, and independent aging desired by older adults [10], [16]–[25]. Technologies such as these offer a way to reduce healthcare costs by facilitating older adults’ ability to age safely at home in less restrictive, less expensive environments. Smart homes can facilitate health and self-care activities by connecting older adults with primary and specialty health care providers, formal home health services, and informal caregivers, to facilitate early interventions and preventions for adverse health events, supporting effective long-term management of chronic conditions while aging in place. These technological solutions also provide an additional layer of safety by continuously monitoring for life and health-threatening situations – in effect, extending the health care workforce.

Domestically, programs such as the MAVHome at the University of Texas Arlington [26], The Aware Home at the Georgia Institute of Technology [27] and the Gator Tech Smart House at the University of Florida [28] have historically served as single-home-test-bed style environments. Internationally, the U-Health smart home project at POSTECH [29]–[31] integrates information from small-sized medical body sensors [32] with other ambient sensors to assist older adults in their homes. Other programs including the Place Lab at the Massachusetts Institute of Technology [33], the Tiger Place project at the University of Missouri-Columbia [34], the CASAS Smart Homes project at Washington State University [35] and the ORCATECH project of the Oregon Health and Science University [36] represent multi-unit smart home projects that are testing a variety of devices such as motion, floor, gait, bed, appliance, temperature, luminance, wearables, smartphone, web-portals, signaling devices, task aids, and other smart/connected devices as a means to impact health and well-being across varying program targets.

1.3 Challenges of Using Ambient Data for Health Monitoring in Private Homes

To date, many such projects are focused on limited user groups (i.e. persons with dementia), institutional settings (i.e. nursing homes), limited sensor types (i.e. contact and/or motion sensors) and/or are narrow in the scope of behaviors they monitor/target (i.e. activities of daily living (ADLs) e.g. bathing and eating, gait, and falls). In addition, the ability to perform long term health trend analysis and detect anomalies in an emergency remains limited.

Using ambient sensor data from private homes to monitor health and wellness is further complicated by the following challenges. Annotated data is lacking due to privacy concerns which preclude the use of cameras, and the well-documented difficulties associated with keeping accurate activity logs in long-term studies. Second, due to various floor plans, furniture arrangements and residents' preferences, ambient networks installed in private homes have significant variation both in terms of the types of sensors used and their deployment. Third, ambient sensor networks are not completely reliable and sensors need periodic maintenance due to malfunction or dead batteries, which results in incomplete observations of the residents' activities. In addition, smart home inhabitants have visitors including family members and friends staying with them from time to time. The activities of visitors trigger the ambient sensors as well as the resident's activities. These changes in the environment, disruptions in the sensor networks, and the presence of visitors introduce data that are not representative of the resident's normal daily activities and lead to an incorrect or incomplete understanding of the residents' activities and wellness.

HomeSense, an ambient health and wellness monitoring platform implemented in community dwellings, provides opportunities to study ambient-sensing solutions of health and wellness monitoring to address the above challenges that are encountered in the real living environments.

1.4 HomeSense: An Ambient Sensing Platform for Health Monitoring

Developed by researchers in the CREATE Health Lab at the University of South Florida, HomeSense is an ambient health and wellness monitoring platform for community-dwelling older adults living independently in their own homes [37]. All participants of HomeSense live alone without pets in their own homes and are recruited from a 55+ active retirement community. Since the start of the study in Aug 2016, 19 participants aged between 68 and 89 have participated for varying lengths of time ranging from 6 to 36+ months. These participants are asked to be available for bi-weekly phone interviews designed to collect self-reported information regarding major health and life events, travel and visitors.

In each participant's home, various wireless sensors are installed to collect information on the participant's daily activities (Figure 1.1). For example, passive infrared (PIR) motion sensors are installed in each room to monitor the occupant's movement within the house, contact sensors attached to medicine boxes, kitchen cabinets and exit/entrance doors to sense the interactions with these items, power sensors attached to electrical household appliances such as coffee pots, washing machines, TVs and microwaves monitor the electricity usage, water sensors detect water usage in toilet tank, and various environmental sensors track changes in temperature, luminance, and humidity in various location in the home. No cameras or microphones are used in any of the deployments.

As noted in Figure 1.2, an array of networked wireless devices are installed in each house and communicate using the Z-wave communication protocol [38]. A Raspberry Pi connected to the Internet acts as a gateway and sends the data from the sensor network to our HIPAA compliant main server using a light-weight machine-to-machine communication protocol MQTT [39]. In the case of Internet connectivity outages, data from the sensors are locally stored and sent to the main server once connectivity is re-established. The sensor data is collected 24/7 and permanently stored in a relational database on the main server. The main server supports communication with the gateways, sensor configuration, device tracking, data visualization, and data analysis activities for the HomeSense project.

Built from the ground up on open source software, since its inception in the field in 2016, HomeSense has had 19 installations over its life course and collected more than 10 million hours of individual, time-stamped sensor data, with an average of 6500 sensor events per day per installation.

1.5 Research Goal and Objectives

Building on the extensive data collected in the HomeSense project, the goal of this research is to develop key indicators and methods to identify pattern changes and disruptions associated with the sensor system itself e.g. sensor or device failures, changes in the environment from which the data is collected e.g. presence of visitors, and changes in the behaviors of the participants which may signal changes in health conditions.

Specific research objectives within this goal are:

1. Develop methods to estimate health and wellness indicators to track changes in overall health and wellness of the residents inhabiting in private homes. Monitoring of these indicators will facilitate the identification of changes that may signal underlying health issues providing actionable information to formal and informal caregivers.

2. Understand and quantify the regularity and predictability of human mobility in private homes. Successful accomplishment of this objective will firmly ground research that builds on regularity and predictability of human activity in a wide range of applications including healthcare, sustainability, and automation. It will also facilitate the development of baseline normal patterns and subsequently the ability to identify changes and disruptions that deviate from these norms.

3. Identify and quantify the impact of different types of changes on the aforementioned indicators by examining records about system changes and the presence of visitors from maintenance logs and bi-weekly assessments. This work will help us understand the characteristics of changes and inform the design of comprehensive metrics capable of detecting a broader range of disruptions in ambient-assisted technologies for health and wellness monitoring.

1.6 Research Work and Contributions

In this section, we briefly summarize the accomplished work for each research objective including the background, the research question, the work that has been completed, experiment results, and the contributions.

1.6.1 Health and Wellness Monitoring Using Ambient Sensor Networks

This work has been published by *Journal of Ambient Intelligence and Smart Environments* [40].

In this study, we present a methodology that estimates occupants' status as active, sedentary, in-bed, out-of-home and unobservable, their location in the house, and their daily activities related to overall health and wellness. The methodology is used to visualize and examine the daily patterns and activities of older adults living in their own homes and participating in a smart home research project. The proposed location and status estimation algorithm is highly accurate as validated by a mobile app that prompts participants with questions about the estimated time of their daily activities. A case study involving a significant health-related life event is presented where the participant's account of changes in her patterns and activities through bi-weekly interviews are shown to confirm inferences based on the results of the proposed methodology.

Details of this work are included in Chapter 2.

1.6.2 An Entropy-based Approach to the Study of Human Mobility and Behavior in Private Homes

This work has been submitted for review by *IEEE Transactions on Human-Machine Systems*.

Understanding human mobility in outdoor environments is critical for many applications including traffic modeling, urban planning, and epidemic modeling. Using data collected from mobile devices, researchers have studied human mobility in outdoor environments and found that human mobility is highly regular and predictable. In this study, we focus on human mobility in private homes. Understanding this type of human mobility is essential as smart-homes and their assistive applications become ubiquitous. We model the movement of a resident using ambient motion sensor data and construct a chronological symbol sequence that represents the resident's movement trajectory. Entropy rate is used to quantify the regularity of the resident's mobility patterns, and an upper bound of predictability is estimated. However, the presence of visitors and malfunctioning sensors result in data that is not representative of the resident's mobility patterns. We apply a change-point detection algorithm based on penalized contrast function to detect these changes, and to identify the time periods when the data does not completely reflect the resident's activities. Experimental results using the data collected from 10 private homes over periods of 178 to 713 days show that human mobility at home is also highly predictable in the range of 70% independent of variations in floor plans and individual daily routines.

Details of this work are included in Chapter 3.

1.6.3 Analysis of Changes in Data Collection Environment in Ambient-assisted Private Homes

Ambient sensor networks are not completely reliable and need periodic maintenance due to malfunction or dead batteries. In addition, smart home inhabitants have visitors including family members and friends staying in their house from time to time. These changes in ambient sensor system and disruptions in the data collection environment introduce data that are not representative of the resident's normal daily activities and may lead to an incorrect or incomplete understanding of the residents' activities and wellness. In this study, we examine and summarize the records about system changes and the presence of visitors from these three sources, i.e., the maintenance log, bi-weekly assessments, and binary information. We construct various metrics to describe the records and analyze the difference between the records that can be corroborated with the change-point detection algorithm introduced in Chapter 3 and those that cannot. Experimental results indicate that the changes in daily entropy rate explain partial changes in sensor systems and the presence of visitors. The records that result in significant changes in the probability distribution of sensor events are more likely to be detected by changes in daily entropy rate.

Details of this work are included in Chapter 4.

1.7 Intellectual Merit and Broader Impacts

In this dissertation, we provide in-depth analyses of a one-of-a-kind dataset that entails thousands of hours of activity data from ambient sensors in 10 private homes. We developed and validated two novel approaches to analyze this type of data; the first one to extract health-related information from ambient sensor data, and the second one to automatically detect disruptions in the sensor systems and the environment from which data is collected. Also in this work, we present a detailed treatment of regularity and predictability of human mobility in private homes. To our knowledge, this is the first time such an analysis has been conducted using data from private homes.

The algorithms developed and tested in this dissertation are critical to the development of an effective ambient-sensing based health and wellness monitoring solutions to enable older adults to age in place. This is desired by both the older adults who wish to age in their own homes and the society that is facing tremendous challenges to maintain sustainable health care services. Smart homes provide information about older adults' daily activities and life routines which can be shared with older adults to encourage them to lead a healthier lifestyle. Based on the collected ambient sensor data, adverse event detection such as fall detection, reminder assistance services such as medicine reminding, and context-aware intelligent services such as energy management, and device automation are developed to provide assistance in the older adults' daily life to improve their quality of life. Information from such systems also provides opportunities to physicians and caregivers to assess and examine the older adults' health status continuously, facilitating early diagnosis and intervention. Overall, the low-cost ambient-sensing smart home technologies provide solutions to allow older adults to live independently in their own homes while aging without reducing their quality of life and reducing health care costs and social burden.

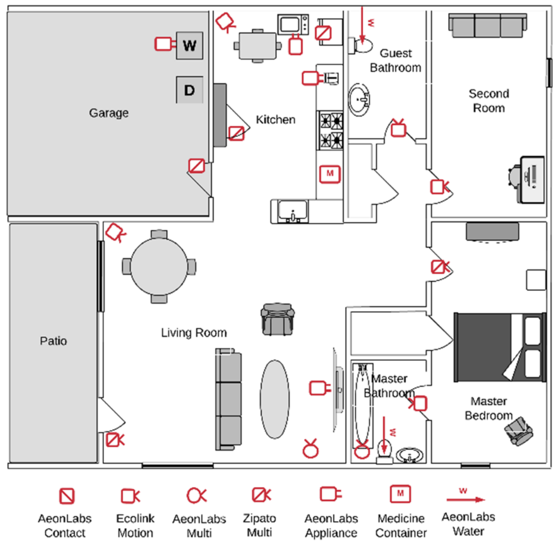


Figure 1.1 Sensor layout for a typical home in HomeSense.

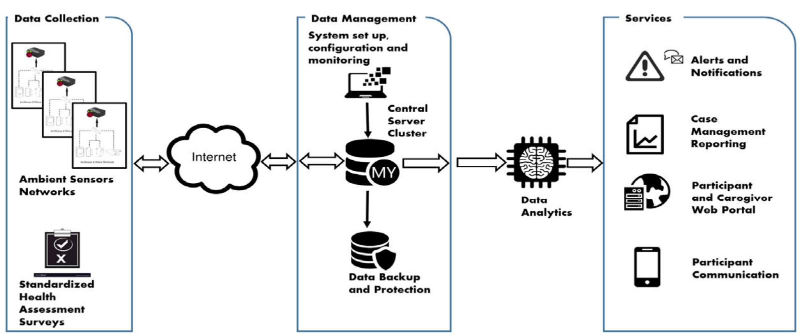


Figure 1.2 HomeSense system overview.

Chapter 2: Health and Wellness Monitoring Using Ambient Sensor Networks

2.1 Note to Reader

Chapter 2 has been previously published by *Journal of Ambient Intelligence and Smart Environments* [40] and has been reproduced with permission from IOS Press. The final publication is available at IOS Press through http://dx.doi.org/ 10.3233/AIS-200553.

2.2 Introduction

Longer life expectancy and aging baby-boomers are causing unprecedented shifts in the U.S. population demographics. The number of people aged 65 and older is projected to reach 83.7 million, almost double that of 2012, making up 21% of the total population in 2050 [41]. This shift towards a predominantly older population is increasing the demand for home and health care services [41], leading to a shortage of skilled workers to provide these services [42], and increasing the informal caregiver burden on the society as a whole [43], [44].

Smart homes with ambient wireless sensors present great potential to manage the growing health care service needs of this older population segment, to improve their quality of life and to enable them to age-in-place [45]–[48], which an overwhelming majority of older adults desire [49]. Smart home projects such as MavHome [50], the Aware Home [51], the PlaceLab [33], CASAS [52], TigerPlace [53], ORCATECH [54], and CASALA [55] are notable research efforts in this direction. Researchers have been able to estimate health and wellness indicators such as movement levels [56], [58]–[61], time outside the home [58], [66], [74], [76], [78] sleep measurements [68]–[70], [72]–[78], walking speed [79] and sedentary activity [80]. The common conclusion among these research efforts is that ambient sensor networks are an effective approach to continuously and unobtrusively monitor the health and wellness of older adults, and the data from these sensor networks are effective in measuring specific health indicators.

Monitoring health and wellness using ambient sensor data can be broadly described in two main categories as sensor-data based and activity-based. In sensor-data based methods, information obtained directly from the sensor data, e.g. time spent in specific locations in the house, frequency of motion sensor firings, is used to detect changes in daily patterns, routines, and overall health and wellness status. Examples of such methods include; in [56] movement patterns based on time spent in specific rooms are used to detect deviations which imply unusual daily life events; in [57] the time spent in rooms and frequency of motion sensor firings are used to model the resident's circadian behaviors, and deviations from this model are used as indicators of anomalies; in [58] the frequency of motion sensor firings per hour is used to construct the density maps to represent the resident's movement patterns over time. The dissimilarity between two density maps are used to measure changes in movement patterns which may indicate health problems; and similarly, in [59] the density maps of the time spent instead of movement levels are used to track changes in the older adult's activity and sleep patterns; and in [60], [61] human stigmergy [62] is used to model spatial-temporal evolution of the resident's movements within the house. The deviations from the reference maps indicate potential changes in the resident's activity routines; in [63] motion sensor data is linguistically summarized to assist clinicians in determining the overall wellness of the older adults.

Activity-based methods, in contrast, focus on estimating health-related activities mostly associated with Activities of Daily Living (ADLs) [64], [65] domains such as outings, sleep patterns, feeding, toileting, and personal hygiene. Notable examples of such work include; in [66] time out-of-home is estimated by a logistic regression model to assess the loneliness of the older adults; in [67] Markov models and a naïve Bayes classifier are used to model activities such as telephone use, hand washing, meal preparation, eating and medication use, and cleaning. These learned models are further used to evaluate the completeness and consistency of daily activities to determine older adults' capability to live independently; in [68]–[70] an ADLs recognition algorithm [71] is utilized and aggregate statistics of the duration to perform these recognized activities and machine learning is used to predict clinical assessment scores, functional health scores, and symptoms relating to Alzheimer's Disease; and in [72] a visual analytics tool is developed to identify abnormal activities based on these recognized activities; in [73] sleep behavior is modeled as a finite state machine with states defined as awake in the bed, asleep in the bed, or out of the bed. The transitions of the state machine characterize sleep behaviors such as bedtime, rise time, sleep latency, and time up at night; in [74] a location tracking algorithm is used to estimate bedtime and rise time based on movement levels and the status of light sensors, and a rule-based approach is used to estimate time out-of-home; in [75] a rule-based model is used to estimate bedtime, rise time and time in bed; in [76] rule-based models are used to estimate time out-of-home and sleep durations to examine older adults' social isolation; in [77] a Bayesian switch-point model is used to identify sleep and wake periods; and in [78] RNNs are used to encode daily activities such as leave home, go to bed, prepare breakfast, use toilet, etc. The trained model is then used to detect deviations from normal daily routines to identify cognitive decline.

In this paper, we present a combined approach to the use of the data collected from ambient wireless sensor networks using both sensor-data based and activity-based methods described above. We process the sensor data in phases gradually discovering information about the occupants' patterns and activities. More specifically, we present a two-phase algorithm where the initial phase estimates the location of the occupant inside the home and his/her status as *active*, *sedentary* and *unknown* using data from motion and contact sensors. The second phase of the algorithm further refines the status estimated in the first phase as *in-bed* and *out-of-home* and revises the estimated locations of the occupant. These status categories are selected based on the common metrics of interest reported in the literature as indicators of health and wellness. We show that raw sensor data converted to an information triplet of (Time, Status, Location) provides valuable insight into the daily patterns and activities of older adults and captures changes in their health and wellness. Unique to the approach described in this paper is that we consider refractory periods of motion sensors in estimating the location of the resident and the time spent in each location, and we use a rule-based algorithm built on information extracted from sensor-data based methods to estimate activities.

The rest of the paper is organized as follows: Section 2.3 describes the data collection environment, the proposed location and status estimation algorithm, and the algorithm validation process. Section 2.4 presents and discusses the results of the algorithm validation process and the results of applying the location and status estimation algorithm to ambient sensor network data from private homes. Section 2.5 describes a case study involving a significant health-related event and how the outputs of the proposed algorithm support and verify a participant's own account of changes in her daily patterns and activities. Finally, Section 2.6 presents conclusions, limitations of the work and future research directions.

2.3 Method

2.3.1 Experimental Environment

HomeSense is an ongoing smart home project at the University of South Florida that aims to use ambient home sensing to enable older adults to manage and coordinate their health care services and age-in-place as long as possible [37].

All participants of HomeSense live alone without pets in their own homes and are recruited from a 55+ active retirement community. Since the start of the study in August 2016, 19 participants aged between 68 and 89 have participated for varying lengths of time ranging from 6 to 36+ months. The participants are asked to be available for bi-weekly phone interviews designed to collect self-reported information regarding major health and life events, travel and visitors. Further details regarding participant recruitment, consent and participation are outlined in IRB Protocol PRO 00020982.

A typical sensor network deployed in HomeSense is shown in Figure 1.1. Passive infrared (PIR) motion sensors in each room sense movement, contact sensors attached to medicine box, kitchen cabinet or exit/entrance doors sense opening and closing of these items, power sensors attached to electrical household appliances such as coffee pots, washing machines, TVs and microwaves monitor the electricity usage, water sensors detect water usage in toilet tank, and various environmental sensors track changes in temperature, luminance, and humidity. No cameras or microphones are used in any of the deployments.

The wireless sensors installed in each house communicate through the Z-wave communication protocol [38]. A Raspberry Pi connected to the Internet acts as a gateway and sends the data from the sensor network to our HIPAA compliant main server using a light-weight machine-to-machine communication protocol MQTT [39]. In case of Internet connectivity outages, data from the sensors are locally stored and sent to the main server once connectivity is re-established. The sensor data is permanently stored in a relational database on the main server. The main server supports communication with the gateways, sensor configuration, device tracking, data visualization, and data analysis activities for the HomeSense project.

2.3.2 Data Collection

The data used in this study considers PIR motion sensors and door/window contact sensors which are the most common sensors used in ambient sensor networks. These are binary sensors that report two values as either ON or OFF. PIR motion sensors report ON when a thermal pattern change is detected in the sensor's field of vision. If no thermal pattern change is detected after a refractory period, an OFF value is reported. Depending on the manufacturer and configuration of the sensors, the refractory periods may range from seconds to minutes. In HomeSense, three different types of motion sensors are used with refractory periods varying between 12 seconds and 4 minutes. Contact sensors have two magnetic parts installed on the door and the door frame. When the door is opened, two magnetic parts are separated and the contact sensor reports an ON value. When the door is closed, two magnetic parts come together and the sensor reports an OFF value.

While the variations in home layouts, furnishings, and personal preferences do not allow specific sensor installation procedures, our research team does follow general guidelines for installation of motion and contact sensors. Motion sensors are installed in every room such that the field of their vision covers the majority of the space in the room where the occupant may be active. In the case of open floor plans and spaces large enough to require more than one motion sensor, they are positioned to minimize the overlap of their field of vision. Contact sensors are installed on all doors e.g. front and garage doors which allow entrance to and exit from the home.

2.3.3 Data Preprocessing

Data preprocessing removes redundancies and standardizes binary data values to prepare the dataset for further analysis. A raw data sample from the sensor system prior to preprocessing is shown in Figure 2.1. Data preprocessing involves three steps:

1. The data is grouped by sensor identity and sorted in ascending time.

2. Sequential identical values reported by the same sensor are eliminated, and only the first (earliest) reported data is kept.

3. Reported binary sensor values (0) and (255 or 1) are standardized to values of 0 and 1 which represent the OFF or ON status respectively.

Data preprocessing results in a sequence of alternating 1 (ON) and 0 (OFF) values in ascending time order for all sensors as shown in Figure 2.2.

2.3.4 Location and Status Estimation Algorithm

The objective of the Location and Status Estimation Algorithm (LSEA) is to estimate a participant's status and location in the house based on the sequence of sensor events. The algorithm has two phases. Phase 1 assigns time segments between sensor events three status categories as *active*, *sedentary* and *unknown* at various locations in the house. The second phase further refines the status and location estimates from the first phase to include the *out-of-home* and *in-bed* status and location. The inputs to LSEA are the preprocessed sensor data for the time period of interest and mappings of these sensors to locations in the house where the sensors are installed. The outputs are estimates of time, status and location of the participant for the time period of interest. The LSEA algorithm steps are described next.

In Phases 1, initial location and status are assigned.

Refractory periods of motion sensors and periods during which the participant is not observable by any of the motion sensors create time periods during which the status and location of the participant are not known directly from the sensor data. The rule-based approach in Phase 1 is designed to minimize such time periods using the sequence of events from motion sensors. The rules described in this phase are independent of the house layout and sensor device manufacturer.

- Step 1.1: Select two consecutive ON values. Denote the time of the first ON value reported by Sensor 1 as ON_1, and the second ON value reported by Sensor 2 as ON_2. The time segment between ON_1 and ON_2 is as assigned an initial status and location according to 4 conditional rules based on the OFF values reported between ON_1 and ON_2 as follows:
  - Rule 1: If there is no OFF_1 event between ON_1 and ON_2, the time segment between ON_1 and ON_2 is assigned status *active* and location *the install location of Sensor 1*. (Note that this rule also covers cases where OFF events from sensors other than Sensor 1 are received between ON_1 and ON_2)
  - Rule 2A: If there is an OFF_1 event between ON_1 and ON_2, then the time segment between ON_1 and OFF_1 is assigned status *active* and location *the install location of Sensor 1*.

The status and location of the time segment between OFF_1 and ON_2 are assigned according to the presence and absence of OFF_2 event as described in Rules 2B1 and 2B2.

  - Rule 2B1: If there is no OFF_2 event between OFF_1 and ON_2, then the time segment between OFF_1 and ON_2 is assigned status *unknown* and location *unknown*.

- Rule 2B2: If there is an OFF_2 event between OFF_1 and ON_2, then the time segment between OFF_1 and ON_2 is assigned status *sedentary* and location *the install location of Sensor 2*.

Table 2.1 illustrates the status and location assignment rules in Step 1.1 for time segments between two consecutive ON events.

- Step 1.2: Repeat Step 1.1 for all pairs of ON events where the first ON event is the second ON event of the previous pair.

Phase 1 transforms the preprocessed data to a series of continuous-time segments with a location and an initial status as *active*, *sedentary* and *unknown*. During the time segments with status and location categorized as *unknown*, the participant may be outside of the home, he/she may be in areas of the home that are unobservable by any motion sensor, or the participant may be sedentary such as sitting in a chair or lying in bed. During the time segments with *sedentary* status, the participant may be sitting on a chair or lying in bed. Phase 2 of the algorithm further revises the *sedentary* and *unknown* status assignments.

In Phase 2, the initial location and status assignments from Phase 1 are revised as *out-of-home*, *in-bed* or *unobservable* based on specific rules. The purpose of these rules is to increase the robustness of the LSEA to variations in sensor types with varying refractory periods, variations in home layouts, and variations in sensor installations due to large pieces of furniture or preferences of the participants. Steps 2.1 and 2.2 re-categorize specific *unknown* time segments as *out-of-home* or *sedentary*. Step 2.3 re-categorizes specific *unknown* and *sedentary* time segments as *in-bed.* Finally, Step 2.4 re-categorizes the remaining *unknown* time segments as *unobservable* since these do not satisfy any of the rules.

- Step 2.1: Iterate through assignments from Phase 1. Revise the status and location of the time segments with status *sedentary* and location *Exit/Entrance* as *out-of-home* (both for status and location).
- Step 2.2: Iterate through assignments from Phase 1. Revise time segments with status *unknown* with the same locations before and after them based on the following two rules:
    - Rule 1: If the locations before and after are *Exit/Entrance*, revise the location and status as *out-of-home*.
    - Rule 2: If the locations before and after are not *Exit/Entrance*, revise the location as the location before or after and status as *sedentary*.
- Step 2.3: Iterate through assignments from the previous steps. Revise time segments with status *unknown* or *sedentary* as *in-bed* (both for status and location) if one of the following rules are satisfied:
    - Rule 1: The duration of the time segment with *unknown* status is greater than a user-defined threshold τ and the location of its next time segment is either a *Master bedroom* or *Master bathroom*.
    - Rule 2: The duration of time segments with *sedentary* status is greater than a user-defined threshold τ and its location is either a *Master bedroom* or *Master bathroom*.

Based on our observations of ambient sensor data from a variety of participants in HomeSense, long time durations with status *unknown* before a motion in *Master bedroom* or *Master bathroom* and long time durations with status *sedentary* and location *Master bedroom* or *Master bathroom* often occur during the night where the participant briefly gets up to go to the bathroom or other part of the house and proceeds to go back to bed. These two rules are designed to categorize these cases as *in-bed* using a parameter τ which allows us to adjust the sensitivity of these rules. In our experience, a value of τ = 20 minutes works well across many homes and the LSEA outputs are robust to any τ value in the range of 10 to 30 minutes.

- Step 2.4: The remaining time segments with status and location *unknown* which cannot be classified in the previous steps are revised as *unobservable* both for status and location.

2.3.5 Illustrative Example

To illustrate the LSEA, consider the preprocessed data in Figure 2.3. The locations of the motion sensors in the house are shown in Table 2.2.

The output of Phase 1 using the preprocessed data of Figure 2.3 is shown in Figure 2.4. Between the first pair of ON events in rows 1 and 2, there is no OFF event pertaining to Sensor 1 (sensor ID = 618). This condition satisfies Rule 1 so the time segment between these two ON events is assigned status *active* and location *Living room* as shown in row 1 in Figure 2.4. The same reasoning categorizes the time segment between the next pair of ON events in rows 2 and 3 in Figure 2.3 as status *active* and location *Living room* in row 2 in Figure 2.4. The next iteration considers ON events in rows 3 and 7 in Figure 2.3. Since there is an OFF event for Sensor 1 (sensor ID = 568) in row 5 which occurs between these two ON events, based on Rule 2A the time segment between rows 3 and 5 is assigned status *active* and location *Exit/Entrance* and the time segment between rows 5 and 7 is assigned status and location *unknown* as shown in rows 3 and 4 in Figure 2.4. For the pair of ON events in rows 13 and 17 in Figure 2.3, there is an OFF event for Sensor 1 (sensor ID = 553) in row 15, and an OFF event for Sensor 2 (sensor ID = 565) in row 16. Based on Rule 2B2, the time segment between row 13 and row 15 is assigned status *active* and location

Master bathroom in row 10 and the time segment between row 15 and row 17 is assigned status *sedentary* and location *Master bedroom* in row 11 in Figure 2.4.

Phase 2 refines the time segments with status *unknown* and *sedentary* from Phase 1 and the output is shown in Figure 2.5. Based on Rule 1 of Step 2.2, the time segment with status *unknown* in row 4 of Figure 2.4 is re-categorized as *out-of-home* both for status and location as shown in row 4 in Figure 2.5. Row 11 in Figure 2.4 is the time segment with status *sedentary*, location *Master bedroom*, and duration greater than $\tau = 20$ minutes; row 13 and row 15 are time segments with status *unknown*, duration greater than $\tau = 20$ minutes, and a location *Master bathroom* or *Master bedroom* after them. Therefore, the time segments in row 11, 13 and 15 based on Step 2.3 are all re-categorized as *in-bed* both for status and location in Figure 2.5. Finally, the *unknown* time segment in row 17 in Figure 2.4 is re-categorized as status *sedentary* and location *Master bedroom* based Rule 2 of Step 2.2 in Figure 2.5. The remaining *unknown* time segment cannot be re-categorized using Step 2.1 to Step 2.3 and according to Step 2.4, it is re-categorized as *unobservable* both for status and location as shown in row 6 in Figure 2.5.

2.3.6 Algorithm Validation

In this study, we selected to validate the two most important location and status outputs of the LSEA, *out-of-home* and *in-bed*, both of which start and end with distinct activities of leaving home and returning home, and waking up and going to bed. These four activities happen less frequently than other activities that cause a change in the status categories and therefore are more recallable for the participants.

To validate the estimated times of these four activities we developed a mobile app nicknamed AVA (Activity Validation App). The app sends participants simple questions in the form of push notifications and queries them about when they performed these activities. For example, the question for the occupant going to bed and the possible responses to the question are shown in the screenshot of AVA in Figure 2.6. We have selected a 20-minute range approximately centered on the estimated time of the activity to balance precision and the participant's ability to recall the approximate time of the activity. The estimated time of an activity is considered accurate if the participant responds by choosing the "Yes, I performed the activity …"

Each participant is sent a maximum of two questions per day. The first question arrives at 11:00 am and is about his/her activities between 6:00 pm the previous day and 11:00 am. The second question arrives at 6:00 pm and is about his/her activities between 11:00 am and 6:00 pm. The questions are randomly sequenced and true negative questions are sent to eliminate habituation [81] and social desirability [82] biases. Questions that are not answered the same day are deleted at midnight.

Two questions "Did you leave home between …" and "Did you return home between …" are used to validate *out-of-home* status. When generating the validation questions for the *out-of-home* status, we include the cases where the estimated *out-of-home* status is longer than 20 minutes. Once the appropriate time segment is identified, the start and end times of the segment are used to generate the questions to be sent to the participant.

Two other questions "Did you go to bed between …" and "Did you get out of bed between …" are used to validate *in-bed* status. When generating the validation questions for the *in-bed* status, we focused on *in-bed* status overnight that occur between 6:00 pm and 11:00 am the next day. If the times between segments with *in-bed* status are less than 30 minutes these are considered as common sleep distributions such as visits to the bathroom. After filtering out these sleep disruptions, the beginning of the first *in-bed* status in the sequence and the end of the last *in-bed* status are estimated as the bed-time and wake-up time respectively. In rare cases where multiple such sequences are found, the sequence with the longest duration is chosen to generate the bed-time and wake-up time.

True negative questions were generated and sent to participants by selecting a random time during the appropriate time period and checking to make sure that it did not coincide with the time of any true positive questions.

2.4 Results and Discussion

2.4.1 Validation of Out-of-home and In-bed Status

Between November 9, 2017 and March 31, 2018 a total of 518 questions were sent to the two participants who agreed to use AVA and participate in the validation of LSEA. Participant 10 downloaded the app and used her own smartphone and Participant 13 used a tablet provided by our research group. The summary of responses by the participant and by question type are shown in Table 2.3. For the four true positive questions used to validate the estimates from the LSEA, the overall accuracy rates calculated as the ratio of number of "Yes" responses to the number of "Yes" or "No" responses, ranged between 90% and 98% with no significant variation by question type or participant showing excellent overall accuracy for the LSEA. For the true negative questions sent to the participants, 88% of the responses were "No, I did not perform this activity in this time range" with no significant variation by participant showing strong evidence that there is no social desirability or habituation bias in the validation process.

2.4.2 LSEA Outputs

The outputs of LSEA are an excellent source of information for visualizing daily patterns and activities of the occupant. For example, Figure 2.7 and Figure 2.8 illustrate the time segments and corresponding status estimates for each day over a 3-month period between December 1, 2017 and February 28, 2018 for two participants in the HomeSense project. These visualizations show daily patterns and reveal patterns that persist or change over time.

Figure 2.7 shows that Participant 4 follows a fairly consistent sleep pattern waking up around 7:00 am and going to bed around 10:00 pm. On January 10, February 24 and February 27, the estimates indicate that the participant may have had restless nights as s/he has spent very early morning hours out of bed. The participant also has a tendency to leave his/her house late morning and early afternoon and usually returns home before 5:00 pm on most days. We can also see that the participant was away from home the night of February 16 returning the next day.

The same type of visualization for Participant 8 is shown in Figure 2.8 for the same time period. This participant also has a fairly consistent sleep pattern however note the increased number and duration of sleep disruptions (yellow and green bars breaking up the blue *in-bed* status) compared with Participant 4. This participant typically has two outings in a day. The first one is usually around lunch hours and the other one usually occurs in the afternoon starting around 4:00 pm. On average, this participant spends 25% of her day out of home compared with Participant 4 who spends 11% of the day out of the home. We can also see that this participant was away from home from January 19 to January 21.

Similarly, Figure 2.9 and Figure 2.10 show LSEA outputs for two more recent participants in the HomeSense project over a five-month period between January 1, 2019 and May 31, 2019. Participant 56 whose daily status is shown in Figure 2.9 has a steady job and is out of the house during business hours three to four days of the week which results in a clearly different pattern than those in Figure 2.7 and Figure 2.8. This participant is also on two short holidays over the five-month period. Figure 2.10 depicts the daily activities of Participant 53 which seems similar to the daily activities of other participants with short to medium outings during the day and a slightly higher number of sleep disruptions at night. The interesting point about this figure is that starting early in May, it seems like the participant does not leave the house at all. In fact, during this time period, even though the participant did leave the house, the contact sensor on the garage door was not working. Therefore, all out-of-home time durations were categorized as *sedentary* when the garage door was used to exit the house which is almost always the case with this participant. We discuss sensor malfunctions and other limitations of the LSEA algorithm in the Conclusions section.

2.5 Case Study

During bi-weekly interviews, one of the participants disclosed details of a planned hip replacement surgery on March 21, 2017. The participant scheduled the surgery in January 2017 when her symptoms were becoming unmanageable. After the surgery, the participant stayed in the hospital for four days and returned home on March 25, 2017. The participant indicated she had several visitors before and after the surgery and that she went to physical therapy outside of her home for eight weeks. Starting in June 2017, the participant stopped mentioning her hip surgery during the bi-weekly interviews.

Figure 2.11 shows the daily status estimates of the LSEA for 14 months from September 2016 to October 2017. The visualization provides useful insight into the participant's activities. We can clearly see that the participant was away from home during Christmas Holidays and also at the end of March for her surgery. The increased *active* status estimates immediately after returning home from the surgery align with the presence of visitors/caregivers in the house. Also note that during the first month after surgery the participant's time *out-of-home* is almost nonexistent and gradually increases over time returning to levels comparable to and exceeding those in September 2016. The overnight *sedentary* status estimations during the Christmas

Holidays and at the end of March after her surgery are caused by the presence of visitors. This is one of the limitations associated with a rule-based approach where the rules are designed for a single occupant and the presence of multiple occupants may cause misclassification of the occupant's status.

A comparison of monthly aggregates of status estimates confirms the observations from Figure 2.11. Figure 2.12 illustrates the monthly summary of the estimated time of the *out-of-home* and *active* status of the participant based on the LSEA over the same 14 month period. In Figure 2.12 (a), there is a notable decrease in *out-of-home* time in the first 7 months which aligns with the worsening of the participant's symptoms prior to surgery. In April 2017, immediately following the surgery, the participant has the lowest *out-of-home* time. *Out-of-home* time increases in the following 6 months as the participant presumably regains her mobility.

Unlike *out-of-home* time which changes in accordance with the participant's mobility, quartiles in box plots of *active* time in Figure 2.12 (b) reach surprisingly high levels in April after the surgery. This seems somewhat counter-intuitive considering the participants reduced mobility due to the surgery. Closer examination of *active* time based on specific locations in the house as shown in Figure 2.13. Figure 2.13 (a) reveals that *active* time is increased in the guest room in March and April which corresponds to the presence of visitors following the surgery. The effect of visitors on increased *active* time can also be seen in the month of December due to visitors during the holiday season. Figure 2.13 (b) which shows increased *sedentary* time in the guest room during these months also confirms the increased use of the guest room presumably by the visitors.

2.6 Conclusion and Future Work

In this paper, we introduce a novel methodology to analyze ambient sensor network data that estimates the status and location of older adults living in their private homes. The methodology captures the daily patterns and activities of the occupant related to overall health and wellness. We demonstrate the potential applications of the methodology using data from private homes of older adults participating in an ongoing smart home research study HomeSense and demonstrate that the outputs of the location and status estimation algorithm are effective in visualizing and quantifying daily patterns and activities, and in investigating changes in overall health and wellness.

Compared with other work in this area, our approach focuses on the totality of the data to estimate general information about the occupants' patterns and activities as opposed to directly estimating particular health and wellness-focused metrics. More specifically, raw sensor data is converted to an information triplet (Time, Status, Location) which is subsequently processed by a rule-based algorithm to estimate the occupants' patterns and activities. This methodology also takes into account inherent variability in the sensor data from ambient sensor networks, e.g. refractory periods, and relies only on the basic information reported by commonly available motion sensors in the market. In this manner, the proposed LSEA is designed to be used with a wide range of ambient sensor networks with any mix of motion sensor types.

Our approach is not without limitations. The rule-based approach to estimating the status of the occupant has only been tested in homes with occupants living by themselves. The presence of long-term visitors causes deviations in the estimation of the occupant's status and location. While these deviations are indicative of the presence of visitors, which itself is an important indicator related to the health and wellness of the occupant, estimates of location and status during such times are not representative of the occupant's activities and patterns. Similarly, false sensor readings (rare but still present) or sensors which stop reporting data due to malfunction or dead batteries do present problems for the proposed approach. Particularly in ambient sensor networks installed in private homes, sensor failures may persist for several days or even weeks if access to the premises to correct the problem is not possible. Finally, the *unobservable* status estimation of the proposed approach lacks any information regarding the occupants' status and location. While it may not be possible to completely eliminate, the duration of time estimated as *unobservable* must be carefully managed through adjustments in sensor deployment and diligent sensor maintenance.

Future work will focus on addressing limitations particularly with respect to imputing missing sensor data in the case of sensor failures to improve the LSEA estimates of the occupants' location and status.

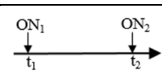
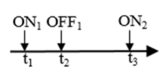
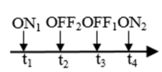
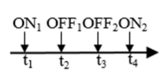
Row	Sensor ID	Time	Value
1	618	2017-01-01 22:09:14.223	1
2	576	2017-01-01 22:09:17.703	255
3	576	2017-01-01 22:09:52.680	0
4	576	2017-01-01 22:10:10.257	0
5	618	2017-01-01 22:13:04.448	0
6	576	2017-01-01 22:36:54.375	255
7	618	2017-01-01 22:36:56.299	1

Figure 2.1 An example of raw sensor data in database server.

Row	Sensor ID	Time	Value
1	618	2017-01-01 22:09:14.223	1
2	576	2017-01-01 22:09:17.703	1
3	576	2017-01-01 22:09:52.680	0
4	618	2017-01-01 22:13:04.448	0
5	576	2017-01-01 22:36:54.375	1
6	618	2017-01-01 22:36:56.299	1

Figure 2.2 The preprocessed form of the raw data in Figure 2.1.

Table 2.1
Location and status estimation algorithm

Rule	Sensor Events	Location and Status Assignment
1		$[t_1, t_2]$ is assigned the *location of Sensor 1* with *active* status.
2A		$[t_1, t_2]$ is assigned *location of Sensor 1* with *active* status; $[t_2, t_3]$ is assigned an *unknown* location with *unknown* status.
2B1		$[t_1, t_3]$ is assigned *location of Sensor 1* with *active* status; $[t_3, t_4]$ is assigned an *unknown* location with *unknown* status.
2B2		$[t_1, t_2]$ is assigned *location of Sensor 1* with *active* status; $[t_2, t_4]$ is assigned *location of Sensor 2* with *sedentary* status.

Row	Sensor ID	Time	Value
1	618	2017-01-01 22:09:14.223	1
2	576	2017-01-01 22:09:17.703	1
3	568	2017-01-01 22:09:51.671	1
4	576	2017-01-01 22:09:52.680	0
5	568	2017-01-01 22:10:05.148	0
6	618	2017-01-01 22:13:04.448	0
7	568	2017-01-01 22:36:48.437	1
8	568	2017-01-01 22:36:52.690	0
9	576	2017-01-01 22:36:54.375	1
10	618	2017-01-01 22:36:56.299	1
11	576	2017-01-01 22:36:57.455	0
12	565	2017-01-01 22:38:18.263	1
13	553	2017-01-01 22:38:39.145	1
14	618	2017-01-01 22:40:00.299	0
15	553	2017-01-01 22:42:57.115	0
16	565	2017-01-01 22:43:02.644	0
17	565	2017-01-02 02:10:22.256	1
18	565	2017-01-02 02:14:26.988	0
19	553	2017-01-02 04:10:59.305	1
20	553	2017-01-02 04:14:36.848	0
21	565	2017-01-02 07:05:53.006	1
22	565	2017-01-02 07:10:15.843	0
23	565	2017-01-02 07:13:20.177	1
24	576	2017-01-02 07:17:31.959	1
25	618	2017-01-02 07:17:36.908	1

Figure 2.3 An example of preprocessed data.

Table 2.2
The mapping of some ambient sensors to locations of a house

Sensor ID	Sensor Type	Install Location
553	Motion sensor	Master bathroom
565	Motion sensor	Master bedroom
568	Door/Window sensor	Exit/Entrance
576	Motion sensor	Living room
618	Motion sensor	Living room

Row	Start Time	End Time	Status	Location
1	2017-01-01 22:09:14.223	2017-01-01 22:09:17.703	*Active*	*Living room*
2	2017-01-01 22:09:17.703	2017-01-01 22:09:51.671	*Active*	*Living room*
3	2017-01-01 22:09:51.671	2017-01-01 22:10:05.148	*Active*	*Exit/Entrance*
4	2017-01-01 22:10:05.148	2017-01-01 22:36:48.437	*Unknown*	*Unknown*
5	2017-01-01 22:36:48.437	2017-01-01 22:36:52.690	*Active*	*Exit/Entrance*
6	2017-01-01 22:36:52.690	2017-01-01 22:36:54.375	*Unknown*	*Unknown*
7	2017-01-01 22:36:54.375	2017-01-01 22:36:56.299	*Active*	*Living room*
8	2017-01-01 22:36:56.299	2017-01-01 22:38:18.263	*Active*	*Living room*
9	2017-01-01 22:38:18.263	2017-01-01 22:38:39.145	*Active*	*Master bedroom*
10	2017-01-01 22:38:39.145	2017-01-01 22:42:57.115	*Active*	*Master bathroom*
11	2017-01-01 22:42:57.115	2017-01-02 02:10:22.256	*Sedentary*	*Master bedroom*
12	2017-01-02 02:10:22.256	2017-01-02 02:14:26.988	*Active*	*Master bedroom*
13	2017-01-02 02:14:26.988	2017-01-02 04:10:59.305	*Unknown*	*Unknown*
14	2017-01-02 04:10:59.305	2017-01-02 04:14:36.848	*Active*	*Master bathroom*
15	2017-01-02 04:14:36.848	2017-01-02 07:05:53.006	*Unknown*	*Unknown*
16	2017-01-02 07:05:53.006	2017-01-02 07:10:15.843	*Active*	*Master bedroom*
17	2017-01-02 07:10:15.843	2017-01-02 07:13:20.177	*Unknown*	*Unknown*
18	2017-01-02 07:13:20.177	2017-01-02 07:17:31.959	*Active*	*Master bedroom*
19	2017-01-02 07:17:31.959	2017-01-02 07:17:36.908	*Active*	*Living room*

Figure 2.4 The output of Phase 1 of the LSEA for the preprocessed data in Figure 2.3.

Row	Start Time	End Time	Status	Location
4	2017-01-01 22:10:05.148	2017-01-01 22:36:48.437	*Out-of-home*	*Out-of-home*
6	2017-01-01 22:36:52.690	2017-01-01 22:36:54.375	*Unobservable*	*Unobservable*
11	2017-01-01 22:42:57.115	2017-01-02 02:10:22.256	*In-bed*	*In-bed*
13	2017-01-02 02:14:26.988	2017-01-02 04:10:59.305	*In-bed*	*In-bed*
15	2017-01-02 04:14:36.848	2017-01-02 07:05:53.006	*In-bed*	*In-bed*
17	2017-01-02 07:10:15.843	2017-01-02 07:13:20.177	*Sedentary*	*Master bedroom*

Figure 2.5 The output of Phase 2 of the LSEA for the preprocessed data in Figure 2.3.

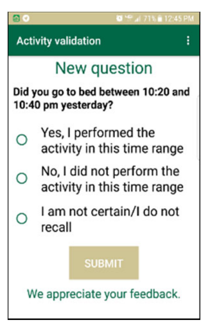


Figure 2.6 An example question of AVA for the activity of ‘going to bed’.

Table 2.3
LSEA validation results

Participant	Question Category	Questions Sent	Response Received	Not Certain	Yes Response	No Response	Accuracy Rate (%)
10	Leave home	45	22	1	20	1	95
	Return home	36	15	4	11	0	100
	Go to bed	30	17	4	12	1	92
	Get out of bed	25	13	0	13	0	100
	True Negative	124	70	3	10	57	85
13	Leave home	45	39	2	32	5	86
	Return home	45	41	3	37	1	97
	Go to bed	26	26	4	20	2	90
	Get out of bed	30	30	2	25	3	89
	True Negative	112	101	2	10	89	90
Overall	Leave home	90	61	3	52	6	90
	Return home	81	56	7	48	1	98
	Go to bed	56	43	8	32	3	91
	Get out of bed	55	43	2	38	3	93
	True Negative	236	171	5	20	146	88

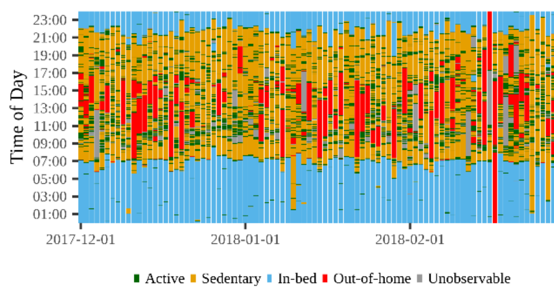


Figure 2.7 Daily estimates of five status categories for Participant 4. The average percentage of daily duration of *active* is 12%, *sedentary* 37%, *in-bed* 36%, *out-of-home* 11%, and *unobservable* 4%.

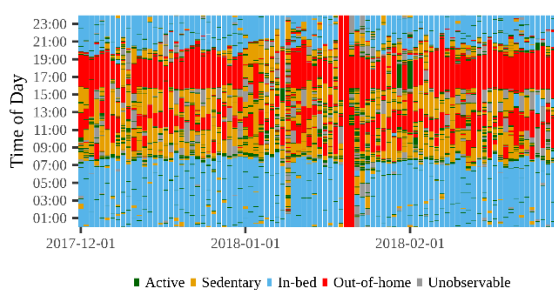


Figure 2.8 Daily estimates of five status categories for Participant 8. The average percentage of daily duration of *active* is 8%, *sedentary* 20%, *in-bed* 41%, *out-of-home* 25%, and *unobservable* 6%.

Figure 2.9 Daily estimates of five status categories for Participant 56. The average percentage of daily duration of *active* is 18%, *sedentary* 11%, *in-bed* 37%, *out-of-home* 31%, and *unobservable* 3%.

Figure 2.10 Daily estimates of five status categories for Participant 53. The average percentage of daily duration of active is 10%, sedentary 40%, in-bed 37%, out-of-home 11%, and unobservable 2%.

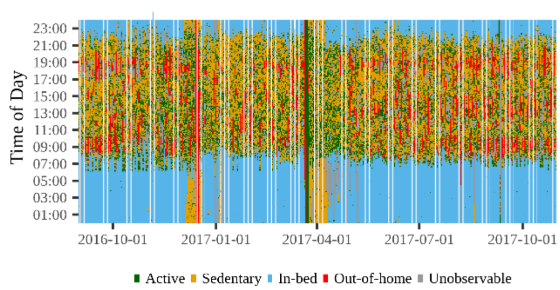


Figure 2.11 Daily estimates of five status categories for Participant 13. The average percentage of daily duration of *active* is 17%, *sedentary* 28%, *in-bed* 35%, *out-of-home* 8%, and *unobservable* 12%.

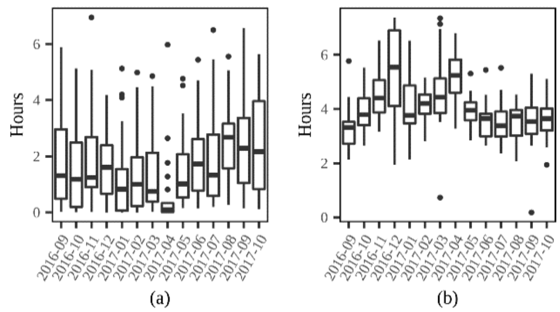


Figure 2.12 Monthly summary of LSEA status estimations for Participant 13 over 14 months (excludes days when the participant was away overnight). (a) *Out-of-home* time. (b) *Active* time.

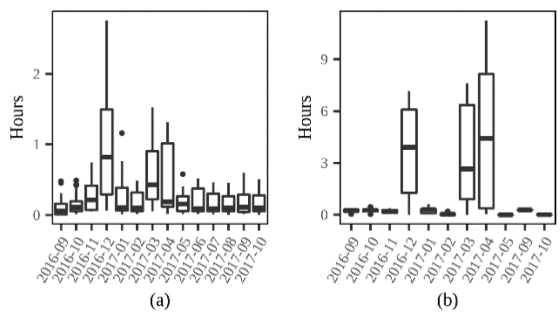


Figure 2.13 Monthly summary of LSEA status and location estimations for Participant 13 over 14 months. (a) *Active* time in the guest room. (b) *Sedentary* time in the guest room.

Chapter 3: An Entropy-based Approach to the Study of Human Mobility and Behavior in Private Homes

3.1 Note to Reader

Portions of Chapter 3 has been submitted for review by *IEEE Transactions on Human-Machine Systems*.

3.2 Introduction

Human mobility is the movement of human beings in space and time and may pertain to an individual or a population [83]. Human mobility occurs in varying distance scales ranging from movement by foot within an indoor environment such as homes or buildings to long-distance travel by different modes of transport using cars, buses, and trains in outdoor environments. In recent decades, the pervasion of mobile devices has enabled the collection of large scale geolocation information related to outdoor human movement facilitating research aimed at gaining a deeper understanding of human mobility. Studies based on ubiquitous data such as call detail records (CDRs) [84], [85], GPS logs [86]–[88], WLAN logs [89], and transportation smart card records [90] have shown that human mobility is not completely random but potentially regular and predictable. Understanding human mobility benefits applications including but not limited to urban planning [91], [92], epidemic models [93], [94], and disaster response [95], [96].

In indoor environments, a growing number of context-aware smart home applications including automation [26], [97], energy management [98], [99], abnormal situation diagnoses [100]–[102], and reminder assistance [103] characterized by their ability to be sensitive to occupants' location, movement, and activity are emerging. Smart homes are increasingly seen as facilitating innovative and supportive environments that provide intelligent services to enable the healthy, safe, and independent aging plan desired by older adults [104], [105]. Domestically, programs such as the MAVHome at the University of Texas Arlington [26], The Aware Home at the Georgia Institute of Technology [27] and the Gator Tech Smart House at the University of Florida [28] have historically served as single-home-test-bed style environments. Internationally, the U-Health smart home project at POSTECH [29]–[31] integrates information from small-sized medical body sensors [32] with other ambient sensors to assist older adults in their homes. Other programs including The Place Lab at the Massachusetts Institute of Technology [33], The Tiger Place project at the University of Missouri-Columbia [34], the CASAS Smart Homes project at Washington State University [35], the ORCATECH project of the Oregon Health and Science University [36], and HomeSense project at the University of South Florida [37] represent multi-unit smart home projects that are testing a variety of devices as a means to impact health and well-being across varying program targets.

The study of human mobility in indoor environments based on ambient sensor data differs from the study of outdoor mobility based on geolocation information in the following five distinct ways.

- *Data collection infrastructure*: In outdoor environments, mobility information is collected through common infrastructures such as mobile communication networks, GPS satellites, Wi-Fi access points, etc. While in indoor environments such as smart homes, the sensor layouts used to collect information differ from house to house due to different floor plans, sensor density and types, and occupant's preferences.

Furthermore, ambient sensors are more prone to temporary outages due to power and usage-related issues resulting in intermittent loss of data.

- *Data generating frequency:* In outdoor environments, data are collected when mobile devices are activated (making a call, accessing some location-related services, or connecting to a Wi-Fi access point), and therefore data generation frequency is sparser than that of ambient sensor networks where sensors are triggered passively without any intent by humans.
- *Data ambiguity:* Mobile devices have unique identifiers linking them to a distinct moving object. On the other hand, data from simple ambient sensors cannot identify one distinct moving object from another. Therefore, visitors and residents in the home would generate a different mobility pattern than only the residents of the home.
- *Distinct location limits:* In outdoor environments, distinct locations humans can visit are essentially unconstrained. However, in smart home environments, the number of distinct locations is fixed and determined by the installed motion sensors.
- *The time period for trajectory construction:* In outdoor environments, an individual's movement over multiple days is modeled as a stationary stochastic process. Typically months of data are needed to capture all visited locations and a single sequence of movements is constructed for each individual in a large population. On the other hand, in smart home environments, a resident repeats routine behaviors on a daily basis. The data collected by ambient sensors facilitates the construction of multiple trajectories for different time periods and enables the study of the changes in human mobility over time.

The design and evaluation of context-aware smart home applications providing adaptive intelligent services for its residents must consider the regularity and predictability of human mobility and behavior at home. The only work we have come across which studies the regularity and predictability of human mobility at home is [106]. In this work, mobility is defined as the number of times an individual moves between different rooms in their home within a specified period of time without explicitly considering location information. The results indicate that while a common model across individuals is absent, a high degree of regularity and predictability of human mobility exists when contextual information e.g. walking speed, age, weather, socioeconomic status, etc. about individuals is taken into consideration. The authors conclude that in-home mobility is also highly stereotyped, albeit in a different way than outdoor mobility, and may have applications in predicting individual human health and functional status by detecting adverse events or trends, and in conducting more meaningful clinical trials.

In this paper, we study human mobility in homes outfitted with ambient sensors. Our objective is to quantify the regularity and predictability of human mobility in private homes. We model an individual's mobility as a stationary stochastic process and construct trajectories of the occupant by sequences of chronologically visited locations using data from ambient motion sensors. The entropy rate of the mobility is estimated from the sequences and represents a quantitative measure of the regularity and the limit of predictability of mobility is estimated using the entropy rate.

The ambiguity associated with the mobility data collected from private homes and the unreliability in the data collection infrastructure introduce significant intermittent deviations to the assumed stationary stochastic process. To capture these unknown number of deviations, we model the time series of daily entropy rate as piecewise constant and estimate these change-points by minimizing a penalized contrast function. [107] and [108] provide comprehensive reviews of methods for change-point estimation in sequential data considering variations in model assumptions. A penalized least-square estimator based on the Schwarz's criterion [109] is introduced in [110] to estimate the unknown number of change-points. In this method, the unknown number of change-points is estimated by minimizing the sum of squares of the residuals combined with a penalty on the number of change-points. It is shown that this least-square estimator is a consistent estimator of the number of change-points under the assumption that the random variables are independent and normally distributed. [111], [112] expanded this work to a general context where the variables are not necessarily independent and proposed to estimate the unknown change-points by minimizing a penalized contrast function which converges to the true values with probability. This method has been used widely in many applications including but not limited to animal trajectory segmentation [113], EEG segmentation [114], CGH data analysis [115], and offset detection in GPS data [116]. In this study, we apply this method to segment the sequence of daily entropy rates to determine change-points.

The rest of the paper is organized as follows: Section 3.3 introduces the theoretical background including the human mobility model, entropy and entropy rate, entropy rate estimation, and the limit of predictability of human mobility; Section 3.4 describes the data collection environment, data preprocessing, trajectory construction, and the dataset used in this study; Section 3.5 describes the methods including change-point detection algorithm, parameter setting, validation of change-points, and an illustrative example; Section 3.6 presents the results of the estimated entropy rate and predictability followed by Section 3.7 where discussion of the results and conclusions are presented.

3.3 Theoretical Background

In this section, we introduce the theoretical fundamentals of human mobility and the background associated with the study of regularity and predictability of human mobility. The notations, definitions, and formulas follow those presented in [117] and [84] where entropy rate has been used to quantify the extent to which an individual's travel patterns are regular and predictable.

3.3.1 Human Mobility Model

Human mobility is modeled as a stationary stochastic process $\boldsymbol{X} = \{X_i\}$, where X_i represents the random variable of the location at time $t_i, i = 1, 2, \dots, n$. Let $\mathcal{X}$ be the set of all possible values of X_i. For a stationary stochastic process, the joint distribution for any subset sequence of random variables in $\{X_i\}$ is invariant for any shift t in time, i.e.,

$$\Pr\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\} = \Pr\{X_{t+1} = x_1, X_{t+2} = x_2, \dots, X_{t+n} = x_n\}$$

In our study, $\mathcal{X}$ is the set of all motion sensors installed in a house, and X_i is a unique motion sensor in this set.

A trajectory is a sample path of $\boldsymbol{X}$ and typically represented as a sequence of time-indexed locations. Let l_i represent the location update at time t_i, a trajectory is then defined as a time series of locations $l_1, l_2, \dots, l_n$ with $t_1 < t_2 < \cdots < t_n$. The duration at l_i is the time difference between t_i and t_{i+1}.

3.3.2 Entropy and Entropy Rate

Let X be a discrete random variable with the probability mass function $p(x) = \Pr(X = x), x \in \mathcal{X}$. *Entropy* of X, denoted as $S(X)$, is defined as

$$S(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x)$$

The unit of entropy is *bit* when the log is to base 2.

Entropy measures the uncertainty of a single random variable. For a random variable with only one possible value, entropy equals 0 indicating there is no uncertainty with the realization of this random variable; while for a random variable with n $(n \neq 0)$ possible values which follows a uniform distribution, entropy equals $\log n$. Generally, a lower entropy implies lower uncertainty in the realization of a random variable.

For a stochastic process $\boldsymbol{X} = \{X_i\}, i \geq 0$, i.e., a collection of random variables indexed by i, entropy rate is defined as

$$S(\boldsymbol{X}) = \lim_{n \to \infty} \frac{1}{n} S(X_1, X_2, \dots, X_n)$$

when the limit exists. In this definition, entropy rate represents the time-averaged entropy of n random variables.

Entropy rate can also be defined as

$$S(\boldsymbol{X}) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} S\left(X_i | X_{i-1},\ X_{i-2},\ \dots,\ X_1\right)$$

when the limit exists. This definition of entropy rate measures the uncertainty of the last random variable given historical information before it. For a stationary stochastic process, both limits exist and are equal [117].

Consider a stochastic process where all random variables are independent and identically distributed. Assume that each random variable is drawn from a uniform distribution with N possible values. The entropy rate of this process is calculated as

$$S^{rand} = \log_2 N \tag{1}$$

which equals the entropy of each random variable. This entropy rate is called *random entropy*. In the study of human mobility, S^{rand} measures the uncertainty of an individual's next location assuming that this individual's movement is totally random among N possible locations.

If the random variables in a stochastic process are independent and follow the same probability distribution $p(i), i = 1, 2, \ldots, N$, where N is the number of all possible locations visited by an individual, the entropy rate of this process is defined as

$$S^{unc} = -\sum_{i=1}^{N} p(i) \log_2 p(i) \tag{2}$$

This entropy rate is referred to as the *temporal-uncorrelated entropy*. Note that all possible locations are equally likely visited when $p(i) = \frac{1}{N}, i = 1, 2, \ldots, N$, which results in $S^{unc} = S^{rand} = \log_2 N$.

The third entropy rate called the *real entropy* and denoted by S^{real}considers the frequency of visited locations and the order in which the locations are visited. It is calculated as

$$S^{real} = -\sum_{T' \in T} P(T') \log_2[P(T')] \tag{3}$$

where T represents the sequence of the visited locations and T' represents a subsequence of T.

Theoretically $S^{real} \leq S^{unc} \leq S^{rand}$. It is important to emphasize that when the process is totally random, $S^{rand} = S^{unc} = S^{real}$, and when the process is not completely random but includes inherent repetitive patterns, S^{real} is the smallest among the three entropy measures.

3.3.3 Entropy Rate Estimation

Given a sequence of length n with N distinct symbols in the sequence, the value of S^{rand} is calculated using (1). To calculate S^{unc} using (2), we need to estimate the probability distribution from the sequence. The probability of $x_i, i = 1, 2, \ldots, N$ is estimated as $\hat{p}(x_i) = N_i/n$, where N_i is the total number of x_i in the sequence. Instead of calculating S^{real} using (3), we estimate the value of S^{real} based on the Burrows-Wheeler block sorting transform (BWT) estimator which is easy to implement and is shown to be almost-sure convergent for stationary, ergodic random processes [118] characteristic of movement trajectories considered in this work.

3.3.4 The Limit of Predictability of Human Mobility

Let $h_{n-1} = \{X_1, X_2, \dots, X_{n-1}\}$ be an individual's locations at times t_1 through t_{n-1} and $P(h_{n-1})$ be the probability of observing h_{n-1}. Let $\pi(h_{n-1})$ be the probability that an individual will be at his/her most likely location at time t_n. The predictability of the $n\ th$ location given the historical trajectory h_{n-1}, denoted as $\Pi(n)$, is defined as

$$\Pi(n) \equiv \sum_{h_{n-1}} P(h_{n-1})\pi(h_{n-1})$$

$\Pi(n)$ can be viewed as the highest accuracy to predict an individual's $n\ th$ location given the historical trajectory h_{n-1}.

Taking the limit, the overall predictability is defined as the averaged predictability over time:

$$\Pi \equiv \lim_{n\to\infty} \frac{1}{n} \sum_{i}^{n} \Pi(i)$$

In [84], the upper bound of predictability Π, denoted as Π^{max}, is obtained by solving the equation

$$S = -\Pi^{max}\log_2(\Pi^{max}) - (1-\Pi^{max})\log_2(1-\Pi^{max}) + (1-\Pi^{max})\log_2(N-1) \qquad (4)$$

where S is the entropy rate and N is the number of distinct symbols in the process. Π^{max} is treated as the theoretical highest accuracy that a best designed predictive algorithm can achieve for the next location prediction problem.

3.4 Data Collection and Preparation

In this section, we introduce the data collection environment, data preprocessing, how trajectories are constructed from ambient sensor data, and the dataset used in this study.

3.4.1 Data Collection Environment

The data used in this study are collected from HomeSense, a smart home project at the University of South Florida that aims to apply ambient intelligence technologies in real living environments to help older adults age in place [37]. Participants of HomeSense are recruited from a retirement community, aged between 55 and 89 and live independently in their own homes. The participants are asked to be available for bi-weekly phone interviews designed to collect self-reported information regarding major health and life events, travel and visitors. Further details regarding participant recruitment, consent and participation are outlined in IRB Protocol PRO 00020982.

A typical sensor network deployed in HomeSense is shown in Figure 1.1. Passive Infrared (PIR) motion sensors in each room sense movement, contact sensors attached to medicine box, pantry, fridge, and exit/entrance doors sense opening and closing of these items, power sensors attached to electrical household appliances such as coffee pots, washing machines, TVs and microwaves monitor the electricity usage, water sensors detect toilet usage, and environmental sensors report changes in temperature, luminance, and humidity. No cameras or microphones are used in any of the deployments.

These wireless sensors installed in each house communicate through the Z-wave communication protocol [38]. A Raspberry Pi connected to the Internet acts as a gateway and sends the data from the sensor network to our HIPAA compliant main server using a light-weight machine-to-machine communication protocol MQTT [39]. In case of Internet connectivity outages, data from the sensors are locally stored and sent to the main server once connectivity is re-established. The sensor data are permanently stored in a relational database on the main server. The main server supports communication with the gateways, sensor configuration, device tracking, data visualization, and data analysis activities for the HomeSense project.

3.4.2 Data Preprocessing

We only use the data from PIR motion sensors for this study. PIRs are installed in every room such that their field of vision covers the majority of the space in the room where the occupant is active. In the case of open floor plans and spaces large enough to require more than one motion sensor, multiple sensors are installed in a way that minimizes the overlap in their fields of vision.

PIRs report two state values as either ON (1 or 255) or OFF (0). PIRs are triggered and report ON when a thermal pattern change is detected in the sensor's field of vision. If no thermal pattern change is detected after a refractory period, an OFF value is reported. The refractory period varies between 12 seconds and 4 minutes depending on the manufacturer of these devices.

The binary sensor data collected from PIR sensors is preprocessed to remove redundancies and standardized before further analysis. Data preprocessing involves three steps: (1) group data by sensor identity and sort them in ascending time; (2) eliminate sequential identical values reported by the same sensor, and only keep the first (earliest) reported data; (3) standardize reported binary sensor values as 0 or 1 which represent status OFF or ON respectively.

3.4.3 Dataset

The dataset includes all participants who were enrolled in HomeSense for at least five months between 2017-01-01 and 2018-12-31, who did not disclose family or friends staying with them long-term, and who did not report significant mental or physical impairments in the bi-weekly assessment. In total 10 homes representing 3812 days of data are used in this study.

Subsequently, 21 more days were excluded from the dataset when the participants reported as being on vacation in bi-weekly interviews, and the days which were not reported by the participants but had fewer than 12 motion sensor events in a given day. This threshold was determined using the sensor data from the days where the participants reported as being on vacation. These events correspond to sensor errors and visitors who may have come to check on the house and are not representative of the participants' typical activities.

3.4.4 Trajectory Construction

Daily motion trajectories are constructed based on the ON events from motion sensors that observe a resident's movement within the house. Consider the motion sensor events in Figure 3.1. An ON event is reported by a motion sensor when a movement is detected in the field of the motion sensor's view, and a sequence of chronological ON events represents the movement history. The sequence of ON events is transferred to a symbol sequence by replacing each of them by the symbol representation, for example, the sensor identity, to uniquely represent the motion sensor that reports an ON event, and thus we construct a symbol sequence representation of movement trajectory of the resident. For the motion sensor data in Figure 3.1 such a trajectory is constructed as '565 – 553 – 553 – 618 – 553 – 618 – 618'.

3.4.5 Data Preparation

Using the daily motion trajectories and the BWT entropy estimator described in Section 3.3.3, we estimate the true daily entropy rate defined in (3) and construct a sequence of daily entropy rates for each home to describe the resident's mobility over time. Similarly, we also calculate the limit of predictability for each day using (4).

We define outliers as data points for which the estimated daily entropy rates are outside of the $[Q_1 - 1.5 * IQR, Q_3 + 1.5 * IQR]$ range where Q_1 and Q_3 are the lower and upper quantile of the dataset respectively, and $IQR = Q_3 - Q_1$. Only outliers that do not have another outlier within ($\pm$3 days) are removed from the dataset to ensure that temporary shifts are not removed from the dataset. Using this method, we exclude 19 data points reducing the dataset size to 3772 for all houses.

Table 3.1 summarizes the resulting dataset size for each house, the minimum and the maximum number of unique symbols in the daily trajectories, and the minimum, the maximum, and the average length of the daily trajectories. The value of the maximum number of unique symbols denoted as N_{max}, varies between 8 and 12 as a consequence of the different sensor layouts in private homes. For houses with the same N_{max}, the average length of daily trajectories also varies from house to house. For example, the average length of the daily trajectory of House 13 (203) is almost twice as that of House 8 (112) while both of them have $N_{max} = 10$, implying that the average movement level of the participant in House 13 is higher than the participant in House 8.

3.5 Methodology

As discussed in the Introduction section, changes in the data collection infrastructure such as addition or removal of sensors, temporary sensor malfunction which may last days or even weeks, and the presence of long-term visitors significantly alter the patterns in the motion sensor data from the residence and the regularity and predictability of the resident's mobility estimated based on it. While such events are unavoidable during longitudinal data collection in private homes, identification and exclusion of such periods of time when the collected data is not truly representative of the resident's normal daily activities will result in a more accurate and representative estimation of the regularity and predictability of the resident's mobility. To accomplish this, we apply a change-point detection algorithm on the sequence of daily entropy rates to identify segments of time where the sensor data may not be representative of the resident's normal activity patterns.

3.5.1 Change-point Detection Algorithm

Denoting the sequence of daily entropy rate as $\boldsymbol{s} = (s_1, s_2, \dots, s_n)$ where n is the number of days, we model this sequence of daily entropy as piecewise constant [119]

$$s_j = \mu_k + \varepsilon_j,\ 0 < \tau_{k-1} < j \le \tau_k < n,\ 1 \le k < K \tag{5}$$

where K is the total number of segments, $\boldsymbol{\tau} = (\tau_1, \tau_2, \dots, \tau_{K-1})$ with $0 < \tau_1 < \tau_2 < \dots < \tau_{K-1} < n$ is the sequence of change-points, μ_k is the mean of daily entropy in segment k which is different for consecutive segments, and ε_j is the error item with a zero mean and a constant variance σ^2.

To estimate the change-points in cases where the true number of change-points are unknown, [108] proposes a model selection via penalization approach where the optimal segmentation solution is obtained by minimizing a penalized contrast function

$$J(\boldsymbol{\tau}, \boldsymbol{s}) + \beta * pen(\boldsymbol{\tau})$$

where $J(\boldsymbol{\tau}, \boldsymbol{s})$ is the contrast function used to measure the contrast between the segmentation marked by $\boldsymbol{\tau}$ and the sequence $\boldsymbol{s}$, $pen(\boldsymbol{\tau})$ is the penalty term which increases as the number of change-points increases, and β is the penalization parameter or tune parameter that adjusts the minimization of $J(\boldsymbol{\tau}, \boldsymbol{s})$ and the minimization of $pen(\boldsymbol{\tau})$. In the rest of this section, we describe the choice of $J(\boldsymbol{\tau}, \boldsymbol{s})$, $pen(\boldsymbol{\tau})$ and an automatic procedure described in [114] to determine the penalized contrast estimate of change-points about the mean of a sequence in (5).

Let $U(s_{\tau_{k-1}+1}, \dots, s_{\tau_k}; \theta)$ be a contrast function to estimate the true value of the attribute $\theta \in \Theta$ of the data points between $s_{\tau_{k-1}+1}$, and s_{τ_k} e.g., mean or variance which doesn't change within the segment k. For any segment, $1 \le k \le K$, the minimized contrast estimation $\hat{\theta}$ satisfies

$$U\big(s_{\tau_{k-1}+1}, \dots, s_{\tau_k}; \hat{\theta}\big) \le U\big(s_{\tau_{k-1}+1}, \dots, s_{\tau_k}; \theta\big), \forall \theta \in \Theta$$

Let

$$G\left(s_{\tau_{k-1}+1}, \dots, s_{\tau_k}\right) = U\left(s_{\tau_{k-1}+1}, \dots, s_{\tau_k}; \hat{\theta}\right)$$

Then, $J(\boldsymbol{\tau}, \boldsymbol{s})$ is defined as the averaged summation of G over all segments:

$$J(\boldsymbol{\tau}, \boldsymbol{s}) = \frac{1}{n}\sum_{k=1}^{K} G\left(s_{\tau_{k-1}+1}, \dots, s_{\tau_k}\right)$$

To estimate the changes in the mean, let

$$U\left(s_{\tau_{k-1}+1}, \dots, s_{\tau_k}; \theta\right) = \sum_{i=\tau_{k-1}+1}^{\tau_k} (s_i - \mu_k)^2$$

Then

$$G\left(s_{\tau_{k-1}+1}, \dots, s_{\tau_k}\right) = \sum_{i=\tau_{k-1}+1}^{\tau_k} (s_i - \bar{s}_{\tau_{k-1}+1:\tau_k})^2$$

where $\bar{s}_{\tau_{k-1}+1:\tau_k} = \frac{1}{n}\sum_{i=\tau_{k-1}+1}^{\tau_k} s_i$, i.e., the estimate of the mean of data in segment $k, 1 \leq k \leq K$.

Thus,

$$J(\boldsymbol{\tau}, \boldsymbol{s}) = \frac{1}{n}\sum_{k=1}^{K} \sum_{i=\tau_{k-1}+1}^{\tau_k} (s_i - \bar{s}_{\tau_{k-1}+1:\tau_k})^2$$

For the penalty function, [114] suggests using the number of segments as the penalty function. Thus $pen(\boldsymbol{\tau}) = K$.

When the number of true segments K is known, the best estimate of $\boldsymbol{\tau}$ denoted as $\hat{\boldsymbol{\tau}}_K$ is the sequence of change-points that minimizes the contrast function $J(\boldsymbol{\tau}, \boldsymbol{s})$. When K is unknown, given an upper bound of K denoted as K_{max}, we can calculate $\hat{\boldsymbol{\tau}}_K$ that minimizes the contrast function for all $K, 1 \dots K_{max}$. By definition, the best choice of K, denoted $\widehat{K}$ among these K_{max} choices is the one that minimizes the summation of the contrast function and the penalty terms $\beta * pen(\boldsymbol{\tau})$.

[114] describes how $\widehat{K}$ varies with the choice of β. For the points in the subset $\{(pen(\boldsymbol{\tau}_{K_i}),\ J(\hat{\boldsymbol{\tau}}_{K_i}, \boldsymbol{s}), i \geq 1\}$ which is the convex hull of the set $\{(pen(\boldsymbol{\tau}_{\boldsymbol{K}}),\ J(\hat{\boldsymbol{\tau}}_K, \boldsymbol{s}), K \geq 1\}$, the value of $\widehat{K}(\beta)$ equals to K_i which remains constant for any $\beta \in (\beta_i, \beta_{i-1})$ where

$$\beta_i = \frac{J(\hat{\boldsymbol{\tau}}_{K_i}, \boldsymbol{s}) - J(\hat{\boldsymbol{\tau}}_{K_{i+1}}, \boldsymbol{s})}{pen(\boldsymbol{\tau}_{K_{i+1}}) - pen(\boldsymbol{\tau}_{K_i})}, i \geq 1$$

[114] suggests choosing the largest $\widehat{K}(\beta)$ for which the length of the interval $[\beta_i, \beta_{i-1}]$ is much larger than that of $[\beta_j, \beta_{j-1}]$ for any $j > i$ as the estimation of the unknown number of segments to capture both significant and minor changes in the sequential data. The process of determining the unknown number of segments described in [114] is summarized as below:

1. For $K = 1, 2, \ldots, K_{max}$, compute $\hat{\boldsymbol{\tau}}_K$ and $J_K = J(\hat{\boldsymbol{\tau}}_K, \boldsymbol{s})$.

2. Compute K_i and β_i for each i and the length (l_i) of the intervals $([\beta_i, \beta_{i-1}])$.

3. Choose the biggest value of K_i such that $l_i \gg l_j$ for $j > i$ as the estimation of the unknown number of segments, i.e., $\hat{\boldsymbol{\tau}}_{K_i}$as the sequence of estimated change-points**.**

We use the following automatic procedure described in [114] to calculate the unknown number of segments.

1. Standardize $J_K = J(\hat{\boldsymbol{\tau}}_K, \boldsymbol{s})$ for any $1 \leq K \leq K_{max}$ by

$$\tilde{J}_K = \frac{J_{K_{max}} - J_K}{J_{K_{max}} - J_1}(K_{max} - 1) + 1$$

where $\tilde{J}_1 = K_{max}$ and $\tilde{J}_{K_{max}} = 1$.

2. For any $2 \leq K \leq K_{max} - 1$, calculate the second derivative $D_K = \tilde{J}_{K-1} - 2\tilde{J}_K + \tilde{J}_{K+1}$. $D_1 = \infty$. Then the minimum penalized contrast estimate of K is

$$\widehat{K} = \max\{1 \leq K \leq K_{max} \text{ such that } D_K \geq S\}$$

where S is a threshold. [114] suggests using $S = 0.75$ based on extensive experimental results.

After determining the number of change-points $\widehat{K}$ and its corresponding segmentation $\hat{\tau}_1, \dots, \hat{\tau}_{\widehat{K}-1}$, we estimate the mean and variance of the daily entropy in each segment using

$$\hat{\mu}_k = \frac{1}{\hat{\tau}_k - \hat{\tau}_{k-1}} \sum_{j=\hat{\tau}_{k-1}+1}^{\hat{\tau}_k} s_j\,, \hat{\tau}_{k-1} < j \le \hat{\tau}_k, 1 \le k \le \widehat{K}$$

$$\hat{\varepsilon}_j = s_j - \hat{\mu}_k, \hat{\tau}_{k-1} < j \le \hat{\tau}_k, 1 \le k \le \widehat{K}$$

3.5.2 Parameter Setting in the Change-point Detection Algorithm

Two parameters are required for the change-point detection algorithm; the minimum number of points in a segment L_{min}, and an upper bound of the number of segments K_{max}.

In our experiment, we use $L_{min} = 1$ to ensure the detection of all possible change-points. For K_{max}, usually a value 2 to 4 times the expected number of segments is suggested to give the algorithm some room to work but to avoid overestimating the number of segments [113], [120], [121]. In our study, the number of changes in the data collection environment and the sensor system, e.g. visitors, sensor system failures, tends to increase as the data collection time period increases. Therefore, longer time periods are more likely to have more change-points. In our experiments, we use the number of weeks contained in the sequential data as the value of K_{max}.

3.5.3 Validation of Change-points

We validate the results of the change-point detection algorithm by checking whether the date of a change-point can be corroborated with the information from three sources; namely the bi-weekly assessments, the maintenance logs, and device battery information collected from the sensor network. We only consider information dated within two days of a change-point as corroborating evidence.

Bi-weekly assessments include information regarding long-term visitors from the participants. In most cases, this information pertains only to visitors who stay with the participant multiple days/weeks, and in many cases the start and end dates of the visit are approximations.

Maintenance logs are used to record the team's maintenance work on the sensor network. Logged maintenance activities include replacement of malfunctioning sensors, repositioning sensors, adding and removing sensors, and replacing batteries all of which impact the observed data. In most cases, to minimize the interruptions to the participants' daily lives, multiple maintenance operations, such as adjusting sensors and replacing batteries, are completed during the same visit.

The third source of information is the data collected from individual devices regarding their battery levels. We use this information to schedule maintenance visits to replace batteries before they are completely drained. If battery replacement is not completed in time and the batteries are completely drained, the device stops reporting data. In such cases, the observed data from the residence, and subsequently the estimates of entropy rates, are not representative of the resident's normal activity patterns.

The validation process entails using the corroborating information from the three sources for the start date of each segment to classify it into one of five categories: (1) *Normal Operation* when the sensor network is completely functional and system is observing only the participant's activities; (2) *System malfunction* when one or more motion sensors malfunction and fail to report data including drained batteries; (3) S*ystem change* when additional motion sensors are added to the system creating a new mode of normal operation; (4) *Visitor presence* when long-term visitors are present, and (5) *Unknown* when we were unable to find corroborating information from bi-weekly assessments or maintenance logs to describe the segment.

3.5.4 Illustrative Example

We use House 55 as an example to illustrate the application of the change-point-detection algorithm introduced in Section 3.5.1 on the sequence of daily entropy rates, and the validation of the detected change-points. The dataset for House 55 has 30 weeks of data. Thus we set the algorithm parameters as $K_{max} = 30, L_{min} = 1$. Figure 3.2 shows the value of the contrast function J_K for $1 \leq K \leq 30$.

The values of K and their corresponding second derivative D_K and J_K for the data points in the convex hull set of (K, J_K) is shown in Table 3.2. Using the automatic procedure and the threshold 0.75, we determine the largest K for which D_K is larger than the threshold 0.75 as the optimal number of segments, i.e., $K = 5$. Figure 3.3 illustrates the five segments of the sequence of daily entropy rates.

Table 3.3 shows the segments and the results of the change-point validation process used to categorize each of the segments. The first segment which covers the dates between 2018-06-05 and 2017-07-06 is categorized as 'Normal_1' based on our best judgment of the system state at that date using the totality of information from bi-weekly assessments and maintenance logs. This categorization is not based on the change-point detection algorithm as the starting point for this segment is the starting date of the dataset. For the second segment, there is no corroborating information for the change-point found at its start date, and thus it is categorized as 'Unknown'. The start date of the third segment 2018-08-08 coincides with a maintenance visit where corrections were made to sensors that were not reporting data and therefore this segment is categorized as 'Normal_1'. The start date of the fourth segment coincides with visitor arrival and the segment is categorized as 'Visitor-related'. The start date of the fifth segment could not be corroborated with any record in the maintenance logs and bi-weekly assessments and therefore this segment is categorized as ‘Unknown’.

This systematic approach to categorizing segments revealed interesting points of change, where the start of a number of ‘Unknown’ segments related to changes in the resident’s life patterns and marked behavioral changes. For example, compared with the fourth segment, sensor events reported by the motion sensors installed in the master bedroom and master bathroom were absent in early mornings starting on 2018-12-18. While this change in the motion sensor events could not be captured by the bi-weekly phone interviews or the maintenance logs, it is caused by the changes in the occupant’s behaviors which explain the change characterizing the fifth segment.

We observe in Figure 3.3 that the mean of the entropy rate and the mean of the limit of predictability changes in successive segments. The mean of the daily entropy rates decreases from 1.48 in Segment 1 to 1.18 in Segment 2 due to system malfunction with the predictability increasing from 0.74 to 0.80; and the mean of daily entropy rates increases from 1.46 in segment 3 to 1.82 in segment 4 due to visitors’ activities with the predictability decreasing from 0.75 to 0.67. We use Welch’s t-test [122] to determine if these differences are statistically significant. The p-values of the t-test for pairwise comparisons of the segments in Figure 3.3 are shown in Table 3.4. We observe that the pairwise comparisons between the mean daily entropy rates and predictability of normal segments are significantly different at the 0.01 level than those of system-malfunction and visitor-related categories, and the results are mixed in the comparisons with the ‘Unknown’ category.

3.6 Results

3.6.1 Overall Entropy Rate and Limit of Predictability

Table 3.5 shows the sample mean, the range of the random, temporal-uncorrelated, and true daily entropy rates over days, and the corresponding limits of predictability for each house. For the entropy measures, the sample mean of the real entropy $\bar{S}^{real}$ is lower than the mean of the temporal-uncorrelated entropy $\bar{S}^{unc}$ and the mean of the random entropy $\bar{S}^{rand}$, providing evidence that there are inherent repetitive patterns in the daily trajectories of the residents. Similar observations are made for the limit of predictability but with a reverse relationship where the mean of the limit of predictability for the real entropy $\overline{\Pi}^{real}$ is the highest. Overall, the sample mean of the real entropy is between 0.48 and 2.36 with a mean of 1.60, and the corresponding limit of predictability is between 54% and 92% with a mean of 72%. The distribution of daily entropy rates and the corresponding limits of predictability for all houses are illustrated by box-plots in Figure 3.4.

3.6.2 Results from the Change-point Detection Algorithm

The real entropy rate measures the extent to which movement patterns are regular. Changes in the regular movement patterns that are caused by changes in sensor system configuration or the visitors' activities could introduce changes in the value of the real entropy rate. The results in this subsection pertain to the analysis of the sequence of daily real entropy rate for each house and use the change-point detection algorithm to examine how it changes over time.

Table 3.6 shows the segments determined by the change-points obtained by the change-point detection algorithm described in Section 3.5.1, and the segment categorizations using the validation process described in Section 3.5.3. There are 37 change-points that are detected over 10 houses where 9 change-points are explained using the visitor-related information in the bi-weekly assessments and 13 are explained using the information in the maintenance logs. In total, 22 out of 37 change-points can be validated by the records of bi-weekly assessment and the maintenance log.

Table 3.7 below summarizes aggregate statistics by segment type from all homes. Note that around 50% of the segments containing 75% of the days correspond to normal behavior. 'Visitor-related' and 'System-malfunction' type segments correspond to around 20% of the segments and less than 10% of the days. 'Normal' type segments are clearly longer containing a significantly higher number of days than those that correspond to visitors and system malfunction. 30% of segments which contain 20% of the days were categorized as unknown. Another observation related to the results in Figure 3.7 is that the range of daily entropy rate of 'Normal' segments (0.81, 2.22) is much narrower than the range of all segments (0.48, 2.36) indicating that those days with uncharacteristically small and large daily entropy rates were not representative of the residents' normal routines, but were associated with disruptions which involved presence of visitors or problems with the ambient sensor system.

3.6.3 Comparison of Entropy Rates between Segment Types

We compare the mean of daily entropy rate of different types of segments within each house to see if there are statistically significant differences between entropy rates of these segments. The results of the 99 pairs of comparisons using Welch's t-test are summarized in Table 3.8. All 'Visitor-related' segments have significantly different means from the 'Normal' segments and all nine 'System-malfunction' segments have significantly different means from the normal ones. As expected, the comparison of means with 'Unknown' segments has mixed results.

3.6.4 Analysis of Normal Days' Entropy Rates

After isolating "Normal" segment types that capture the routine behavior of the residents, we proceeded to compare entropy rates of these days within and across homes to determine if there are meaningful subgroups or trends. We first compared entropy rates of weekdays with weekends as shown in the box plot in Figure 3.5. The sample means of daily entropy rates for weekdays and weekends were 1.64 and 1.63 respectively showing no significant difference between the means (p-value = 0.26). For this cohort, we did not expect to see a difference as only one participant has a routine work schedule. This participant works 10 hours each day on Wednesday, Thursday, and Friday and the box plot of daily entropy rate for each day of the week for this participant is shown in Figure 3.6. We did not observe any significant differences in daily entropies between three working days with the non-working days in Figure 3.6. Comparison of the entropy rates of the three working days with the non-working days also did not show any significant differences between the means of daily entropies (p-value = 0.51).

When we studied the daily entropy rates of the participants stratified by age group, we obtained very interesting results. Of the 10 participants, two are below age 70, two are between the ages of 70 and 75, and six are older than 75. The box plots of the entropy rates for these three age groups are shown in Figure 3.7. The sample means are 1.48, 1.55, and 1.67 respectively and show statistically significant differences ($p < 0.001$) in the daily entropy rates among different age cohorts.

3.7 Discussion and Conclusions

In this paper, we studied human mobility in private homes using data from ambient sensors that observe residents' movements. We construct daily movement trajectories based on the collected sensor data and use entropy rate to measure the regularity and predictability of these trajectories. Our analysis shows that the movements of these residents at home are not completely random, but inherently regular and are predictable. The average real entropy for daily trajectories range between 0.81 and 2.22, and their corresponding limit of predictability is between 0.56 and 0.86 (Table 3.7). On average, about 70% of the time the resident's next location can be correctly predicted by a theoretically best designed predictive algorithm. The regularity and predictability of the resident's movements under conditions representative of normal life routines, across different homes with varying floor plans, and for individuals with different lifestyles remained within a very narrow range over long periods of time. This is a very important finding and a unique contribution of this research. To our knowledge, it is the only work of this kind to quantify the predictability of human mobility in private homes and demonstrate its consistency across 10 installations and 3772 days of data.

The data collected from wireless ambient sensor systems in private homes over extended periods of time contains temporary shifts predominantly due to the presence of visitors in the homes and malfunctions in the sensor systems. These factors skew the data collected from the home in the form of missing sensor data in the case of system malfunctions, and additional sensor data not representative of the resident's movements in the case of visitors. A change-point detection algorithm is used to identify such segments of time and study their influence on the entropy rates of daily trajectories. Results of the change-point detection algorithm shown in Tables 3.6 and 3.7 present clear differences between the entropy rates of days that belong to different types of segments.

Using the bi-weekly phone interviews with the participants and maintenance logs to corroborate the change-points from the algorithm, the segments were classified into four categories as 'Normal' ('Normal_1' and 'Normal_2'), 'Visitor-related', 'System-malfunction', and

‘Unknown’. 75% of the study days corresponded to the normal behavior of the participant without the effects of known artifacts such as visitors and sensor system malfunctions. ‘Visitor-related’ and ‘System-malfunction’ type segments corresponded to less than 10% of the days, and 20% of the days were categorized as ‘Unknown’ as the starting change-points could not be validated by the interviews and logs. However, we were able to anecdotally observe behaviors from the rest of the sensor data which could have caused changes in daily entropy rate associated with the behavior of the participant such as changes in sleeping habits which coincided with the start of an unknown period. We note the detection of participants’ behavioral changes using entropy rate as an important future research direction.

‘Normal’ type segments were much longer in duration and contained a significantly higher number of days than those that correspond to visitors and system malfunction. While the average daily entropy rate of the normal days was comparable to the overall average daily entropy rate (1.64 vs 1.60), the range of observed daily entropy values of the normal days was significantly narrower. We also observed consistent and statistically significant differences in the means of daily entropies for days categorized as ‘Normal’ vs. ‘System-malfunction’ and ‘Visitor-related’ as shown in Table 3.8. The mean daily entropy rate of visitor days was on average higher than days categorized as ‘Normal’ and ‘System-malfunction’. This is somewhat intuitive as during these days the presence of visitors in the house increased the amount of entropy. On the other hand, days during which there were sensor malfunctions where one or more sensors failed to send data, the average daily entropy rate was lower.

After isolating the effect of known causes on the daily entropy rate and focusing on days categorized as ‘Normal’ segments, we proceeded to analyze the data across homes to identify potential patterns. Since our participants are retired older adults, we did not observe any significant differences in daily entropies between weekdays and weekends as shown in Figure 3.6. Analysis of the daily entropies of the days of the week for one of our participants who works a regular schedule three days a week also did not show significant differences in daily entropy. While this is a very small dataset, it does provide additional evidence that an entropy-based approach is robust to varying lifestyles and routines.

The most interesting results were obtained when analyzing daily entropy rate stratified by age group. We observed statistically significant increases in average daily entropy rate for older cohorts as shown in Figure 3.7. While our dataset is small based on 10 participants, this is a novel and interesting finding which motivates the further study of entropy-based metrics that measure the amount of disorder in stochastic processes as part of an ambient home monitoring system to identify aging-related behavior changes.

Overall, 60% of the change-points detected by the algorithm are validated by the information in the bi-weekly phone interviews with the participants and maintenance and system logs. Since the information from the logs are incomplete, and there were other potential sources of change in the data collected from the private homes such as the changes in the residents behavior, we believe this percentage of validation is in fact very promising in terms of further investigating entropy-based metrics as part of a comprehensive activity and overall health monitoring system in more structured and closely monitored experimental designs. Identification of periods of time which are skewed by factors other than participants' behaviors is essential for effective monitoring of health and wellness using ambient sensor systems in private homes.

Sensor ID	Time	State
565	06:59:47	**ON**
553	07:06:34	**ON**
553	07:12:15	OFF
553	07:16:56	**ON**
553	07:21:32	OFF
565	07:22:53	OFF
618	07:39:21	**ON**
553	07:52:33	**ON**
618	07:54:42	OFF
618	08:01:17	**ON**
618	08:03:56	OFF
618	08:08:09	**ON**

Figure 3.1 An example of dataset.

Table 3.1
A summary of datasets for 10 houses

House	Size of datasets after removing outliers	N_{min}	N_{max}	Minimum trajectory length	Maximum trajectory length	Averaged trajectory length
8	687	4	10	21	395	112
13	713	4	10	23	545	203
14	178	3	8	28	181	82
27	674	3	8	19	264	96
28	495	6	11	17	542	192
51	178	5	10	15	286	131
53	210	5	9	31	197	92
54	220	6	12	50	492	212
55	208	4	10	37	368	168
56	209	5	10	38	529	173

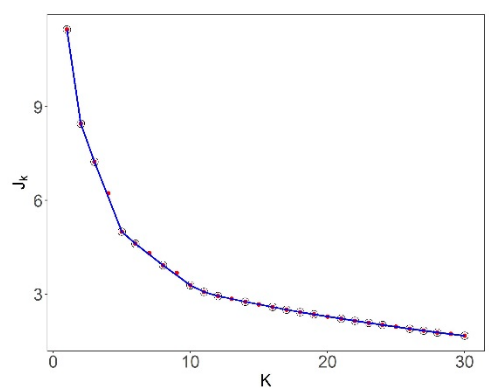


Figure 3.2 The value of the contrast function J_K for $1 \leq K \leq K_{max} = 30$ for House 55. Circles indicate the convex hull points of (K, J_K).

Table 3.2

The values of the second derivative D_K and contrast function J_K for the convex hull points in (K, J_K)

K	D_K	J_K
1	Infinity	11.47
2	5.25	8.46
3	0.68	7.23
5	2.50	5.00
6	0.25	4.61
8	0.41	3.93
10	0.53	3.29
11	0.25	3.07
12	0.11	2.94

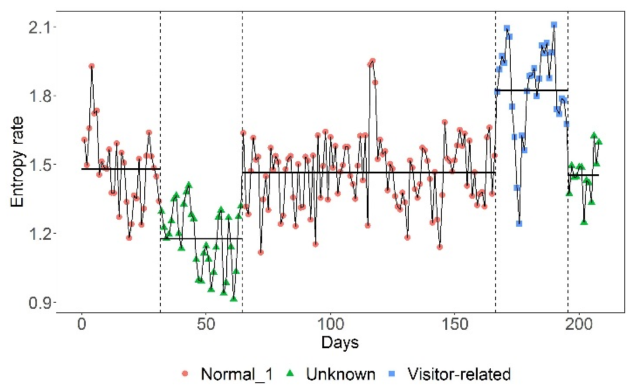


Figure 3.3 The daily entropy rates in five segments for House 55. The black horizontal lines in the graph show the sample means of the daily entropy rate for each segment, and the vertical dashed lines indicate the location of four change-points.

Table 3.3

Five segments obtained by the change-point detection algorithm in House 55

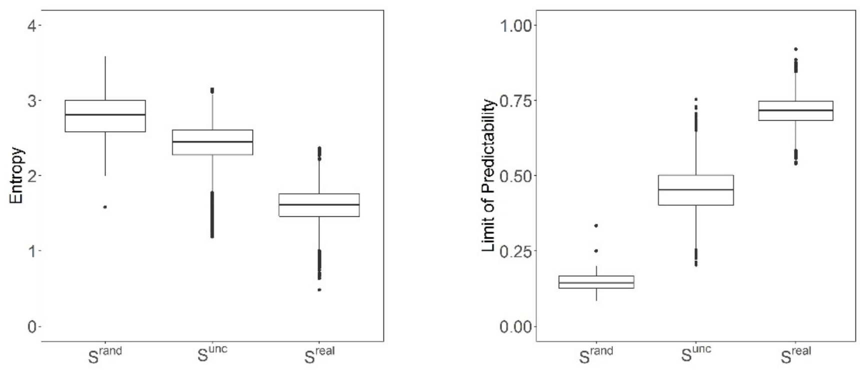
Segment	Number of data points	Date start	$\bar{S}^{real}$ (SD)	$\bar{\Pi}^{real}$ (SD)	Interpretation of the start date	Segment type
1	31	2018-06-05	1.48 (0.16)	0.74 (0.036)	Not applicable	Normal_1
2	33	2018-07-06	1.18 (0.14)	0.80 (0.026)	Unknown	Unknown
3	102	2018-08-08	1.46 (0.15)	0.75 (0.032)	Replace a malfunction sensor	Normal_1
4	29	2018-11-19	1.82 (0.20)	0.67 (0.047)	Visitor activity	Visitor-related
5	13	2018-12-19	1.45 (0.10)	0.74 (0.030)	Unknown	Unknown

Table 3.4
The p-values of the t-tests of the daily entropy rate (predictability) for pairs of segments in House 55

Segment (Segment type)	Segment 2 (Unknown)	Segment 3 (Normal_1)	Segment 4 (Visitor related)	Segment 5 (Unknown)
Segment 1 (Normal_1)	7.96e-11 (1.73e-08)	0.63 (0.57)	1.51e-09 (1.00e-08)	0.53 (0.42)
Segment 2 (Unknown)	NA	3.23e-14 (1.85e-12)	1.81e-19 (3.17e-16)	2.26e-08 (3.54e-06)
Segment 3 (Normal_1)	-	NA	6.75e-11 (5.38e-10)	0.75 (0.17)
Segment 4 (Visitor related)	-	-	NA	9.65e-10 (5.34e-06)
Segment 5 (Unknown)	-	-	-	NA

Table 3.5
The sample means of entropy rate and the limit of predictability

House	$\bar{S}^{rand}$ $[S^{ramd}_{min}, S^{ramd}_{max}]$	$\bar{S}^{unc}$ $[S^{unc}_{min}, S^{unc}_{max}]$	$\bar{S}^{real}$ $[S^{real}_{min}, S^{real}_{max}]$	$\bar{\prod}^{rand}$ $[\prod^{rand}_{min}, \prod^{rand}_{max}]$	$\bar{\prod}^{unc}$ $[\prod^{unc}_{min}, \prod^{unc}_{max}]$	$\bar{\prod}^{real}$ $[\prod^{real}_{min}, \prod^{real}_{max}]$
8	2.91 [2.00, 3.32]	2.45 [1.86, 2.84]	1.57 [1.14, 2.29]	0.14 [0.10, 0.25]	0.46 [0.33, 0.60]	0.73 [0.54, 0.83]
13	2.92 [2.00, 3.32]	2.65 [1.86, 3.15]	1.82 [1.20, 2.36]	0.14 [0.10, 0.25]	0.37 [0.20, 0.55]	0.67 [0.56, 0.75]
14	2.41 [1.58, 3.00]	2.01 [1.19, 2.64]	1.31 [0.48, 2.01]	0.20 [0.13, 0.33]	0.52 [0.34, 0.73]	0.76 [0.63, 0.92]
27	2.55 [1.58, 3.00]	2.24 [1.28, 2.70]	1.53 [0.70, 2.14]	0.18 [0.13, 0.33]	0.45 [0.26, 0.72]	0.71 [0.59, 0.89]
51	3.09 [2.32, 3.32]	2.35 [1.64, 2.73]	1.58 [1.17, 2.00]	0.12 [0.10, 0.20]	0.53 [0.35, 0.70]	0.74 [0.66, 0.82]
53	2.81 [2.32, 3.17]	2.41 [2.05, 2.76]	1.47 [1.14, 1.88]	0.15 [0.11, 0.20]	0.45 [0.31, 0.58]	0.75 [0.67, 0.82]
54	3.18 [2.58, 3.58]	2.40 [1.44, 2.89]	1.66 [1.00, 2.14]	0.11 [0.084, 0.17]	0.53 [0.29, 0.75]	0.73 [0.61, 0.85]
55	2.78 [2.00, 3.32]	2.33 [1.72, 2.76]	1.47 [0.91, 2.11]	0.15 [0.10, 0.25]	0.48 [0.28, 0.62]	0.74 [0.60, 0.84]
56	2.61 [2.32, 3.32]	2.11 [1.68, 2.53]	1.42 [1.05, 2.12]	0.17 [0.10, 0.20]	0.52 [0.38, 0.67]	0.74 [0.62, 0.82]
Overall	2.85 [1.58, 3.58]	2.41 [1.19, 3.15]	1.60 [0.48, 2.36]	0.14 [0.084, 0.33]	0.45 [0.20, 0.75]	0.72 [0.54, 0.92]



(a) (b)

Figure 3.4 (a) Box plots of three entropy measures for all 10 houses. (b) Box plots of the limit of predictability of three entropy measures for all 10 houses.

Table 3.6
Segments of the sequence of daily entropy rates over 10 houses and the validation results

House	Segment	Number of data points	Start Date	$\bar{S}^{real}$ (SD)	Interpretation of start date	Segment type
	1	22	2017-01-01	1.42 (0.14)	Not applicable	Normal_1
	2	36	2017-01-24	1.89 (0.15)	Visitors arrived	Visitor-related
	3	233	2017-03-01	1.55 (0.14)	Visitors left	Normal_1
	4	11	2017-11-03	1.82 (0.10)	Visitors arrived	Visitor-related
	5	358	2017-11-14	1.54 (0.13)	Visitors left	Normal_1
8	6	27	2018-12-05	1.68 (0.15)	Unknown	Unknown
	1	222	2017-01-01	1.66 (0.13)	Not applicable	Normal_1
	2	180	2017-08-15	1.88 (0.12)	Add a new sensor	Normal_2
	3	13	2018-02-11	2.22 (0.11)	Visitors arrived	Visitor-related
13	4	298	2018-02-25	1.89 (0.12)	Visitors left	Normal_2
	1	38	2017-01-01	1.23 (0.19)	Not applicable	Normal_1
	2	54	2017-02-08	0.93 (0.18)	Unknown	Unknown
14	3	86	2017-04-03	1.59 (0.15)	Lower three sensors' view for a better coverage	Normal_2
	1	21	2017-01-01	1.09 (0.20)	Not applicable	System-malfunction
	2	66	2017-01-25	1.29 (0.16)	Replace a drained battery	Normal_1
	3	126	2017-04-01	1.44 (0.16)	Unknown	Unknown
	4	126	2017-08-15	1.62 (0.16)	Add two new sensors	Normal_2
	5	57	2017-12-21	1.40 (0.16)	Unknown	Unknown
	6	192	2018-03-13	1.69 (0.14)	Replace a drained battery	Normal_2
27	7	86	2018-09-30	1.58 (0.17)	Unknown	Unknown
	1	38	2017-07-07	1.56 (0.14)	Not applicable	Normal_1
	2	114	2017-08-15	1.72 (0.12)	Add a new sensor	Normal_2
	3	8	2017-12-16	2.01 (0.16)	Visitors arrived for Christmas	Visitor-related
	4	175	2017-12-30	1.76 (0.13)	Visitors left	Normal_2
	5	64	2018-07-12	1.62 (0.12)	Unknown	Unknown
	6	6	2018-09-27	2.09 (0.16)	Unknown	Unknown
28	7	90	2018-10-03	1.70 (0.12)	Unknown	Unknown
	1	21	2018-05-14	1.32 (0.11)	Not applicable	System-malfunction
51	2	157	2018-06-11	1.62 (0.14)	Replace a malfunctioned sensor	Normal_1
	1	60	2018-05-23	1.43 (0.17)	Not applicable	Normal_1
	2	53	2018-07-23	1.53 (0.12)	Unknown	Unknown
	3	28	2018-09-26	1.33 (0.12)	Sensor malfunction due to drained battery	System-malfunction
53	4	69	2018-10-24	1.51 (0.14)	Replace two drained batteries	Normal_1
	1	37	2018-05-21	1.61 (0.18)	Not applicable	Normal_1
	2	30	2018-06-27	1.42 (0.17)	Unknown	Unknown
	3	41	2018-07-27	1.64 (0.13)	Unknown	Unknown
	4	16	2018-09-11	1.94 (0.13)	Sensor malfunction due to network issue	System-malfunction
54	5	96	2018-09-27	1.72 (0.13)	Repair dropped sensor	Normal_1
	1	31	2018-06-05	1.48 (0.16)	Not applicable	Normal_1
55	2	33	2018-07-06	1.18 (0.14)	Unknown	Unknown

Table 3.6 (Continued)

	3	102	2018-08-08	1.46 (0.15)	Replace a malfunctioned sensor	Normal_1
	4	29	2018-11-19	1.82 (0.20)	Visitor activity	Visitor-related
55	5	13	2018-12-19	1.45 (0.10)	Unknown	Unknown
	1	66	2018-06-04	1.53 (0.10)	Not applicable	Normal_1
	2	5	2018-08-10	1.96 (0.12)	Unknown	Unknown
	3	22	2018-08-15	1.51 (0.11)	Unknown	Unknown
56	4	116	2018-09-07	1.32 (0.11)	Sensor malfunction due to drained battery	System-malfunction

Table 3.7
Aggregate statistics (mean, (standard deviation) [minimum, maximum]) of daily entropy rate and limit of predictability of different types of segments over 10 houses

Type	Normal_1	Normal_2	Normal_1& Normal_2	Visitor-related	System malfunction	Unknown	Overall
Number of segments	15	7	22	5	5	15	47
Number of days	1595	1171	2766	97	202	707	3772
$\bar{S}^{real}$ (SD) [min, max]	1.55 (0.17) [0.81, 2.10]	1.77 (0.17) [1.16, 2.22]	1.64 (0.20) **[0.81, 2.22]**	1.92 (0.20) [1.24, 2.36]	1.35 (0.22) [0.70, 2.14]	1.49 (0.26) [0.48, 2.34]	1.60 (0.24) **[0.48, 2.36]**
$\bar{\Pi}^{real}$ (SD) [min, max]	0.73 (0.040) [0.56, 0.86]	0.69 (0.039) [0.57, 0.82]	0.71 (0.044) [0.56, 0.86]	0.65 (0.045) [0.54, 0.80]	0.76 (0.044) [0.61, 0.89]	0.73 (0.052) [0.57, 0.92]	0.72 (0.048) [0.54, 0.92]

Table 3.8
The number of t-test with p-value < 0.01 vs. the number of t-test with p-value >= 0.01 for comparing the means of entropy rates in two segments

	Normal_2	Visitor-related	System malfunction	Unknown
Normal_1	7 vs. 0	**10 vs. 0**	**7 vs. 0**	16 vs. 6
Normal_2	**-**	**4 vs. 0**	**2 vs. 0**	11 vs. 2
Visitor-related	-	-	-	6 vs. 1
System malfunction	-	-	-	8 vs. 0
Unknown	-	-	-	-

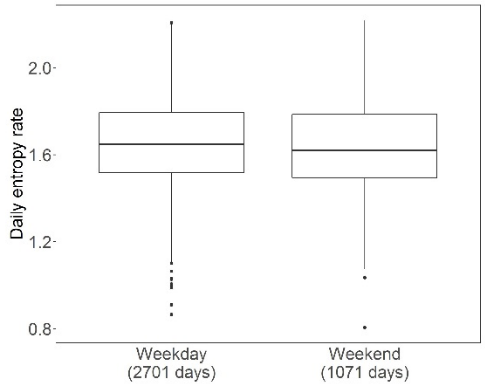


Figure 3.5 Box plots of daily entropy for weekday vs. weekend from normal segments (Normal_1 and Normal_2).

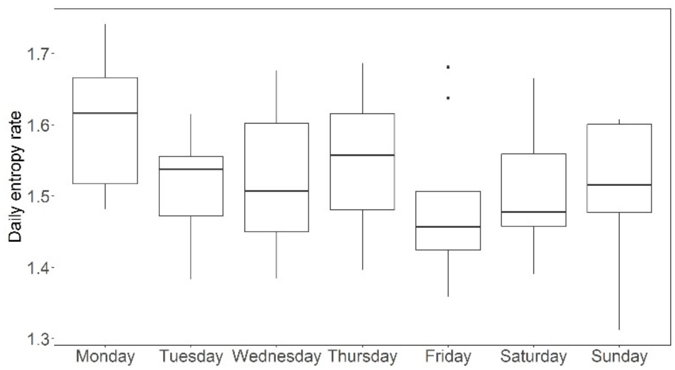


Figure 3.6 Box plots of daily entropy rate for each day of 30 weeks for a participant who works on Wednesdays, Thursdays, and Fridays.

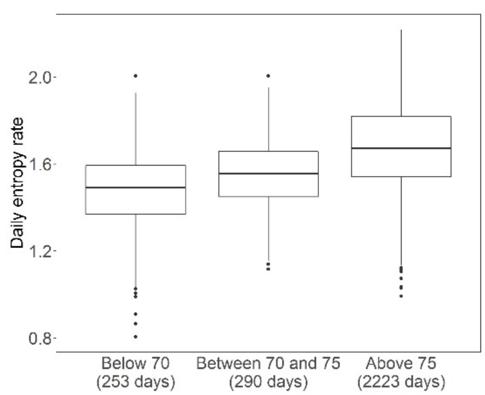


Figure 3.7 Box plots of the real entropy rates for three age cohorts.

Chapter 4: Analysis of Changes in Data Collection Environment in Ambient-assisted Private Homes

4.1 Introduction

HomeSense uses ambient sensor networks to continuously collect data about the residents' daily activities. The collected data is assumed to be representative of the residents' daily activities and is further used for health and wellness monitoring. However, ambient sensor networks are not 100% reliable and sensors in the networks need periodic maintenance due to malfunction or dead batteries, which results in incomplete observation of the residents' activities. In addition, smart home inhabitants have visitors including family members and friends staying in their house from time to time. The activities of visitors trigger the ambient sensors as well as the resident's activities. These changes in the environment and disruptions in the sensor networks and the presence of visitors introduce data that are not representative of the resident's normal daily activities and may lead to an incorrect or incomplete understanding of the residents' activities and wellness.

Recall that in Chapter 3, we apply a change point detection algorithm on the sequence of daily entropy rate to identify time periods when the collected data reflects the resident's daily activities. We validate detected change-points by checking whether the date of the change-points can be corroborated with the information from three sources; namely the bi-weekly assessments, the maintenance logs, and device battery information collected from the sensor network. In this chapter, we work the other way around, which is we examine the records about system changes and the presence of visitors from these three sources, and identify which of them can be corroborated with detected change-points in the sequence of daily entropy rate. To be specific, we divide the records about system changes and the presence of visitors respectively into two groups: one consists of records that can be corroborated with detected change-points, the other consists of records that are not. We construct metrics to describe the records in these two groups and analyze their difference. This work will help us understand the characteristics of changes that are more likely to be detected by changes in daily entropy rate and inform the design of more comprehensive metrics capable of detecting a broader range of disruptions in ambient-assisted technologies for health and wellness monitoring.

The rest of this chapter is organized as follows: In Section 4.2 we provide an overview of records about system changes (we call them 'change records' for convenience) in the maintenance log and device battery information. Then in Section 4.3, we design metrics to depict and compare characteristics of change records in two groups. After that, we summarize the records about the presence of visitors (we call them visitor records) from bi-weekly assessments and examine their characteristics in Section 4.4 and Section 4.5 respectively. Finally, in Section 4.6, we summarize our findings.

4.2 Overview of Change Records in Maintenance Log and Device Battery Information

The records about changes in sensor systems and device battery information over 10 houses are summarized in Table 4.1. In total there are 42 records about changes in sensor systems including adding new motion sensors, removing motion sensors from the current sensor system, adjusting the field of the motion sensors' view for a better coverage, reinstalling motion sensors to the original location if they fall off from the wall, relocating motion sensors to a new location, replacing the dead batteries with new ones, and replacing the malfunctioned sensors with sensors that function well.

For example, according to the information from columns in Table 4.1, we know that the first record occurred in House 8 and it is about changing the location of the motion sensor with the alias 'In Bathroom, Guest' for better coverage on 2017-01-24. The second record is to replace a malfunctioned motion sensor with a good one on 2017-05-17; on the same day, the dead battery of the motion sensor with the alias 'In Shower' was replaced with a new battery. The second and third maintenance occurred on the same day and we treat these two changes as one change. The third record is about the replacement of a dead battery. Noticing that this sensor's battery was dead before the maintenance, we checked the battery sensor events and wakeup sensor events of the sensor 'In Shower' and find the date when both the battery and wakeup sensor failed to report an event as the true date when the battery run out, which generates a new record in row 4 with the description of 'No data reported due to dead battery' on 2017-02-14. A similar situation can be found in rows 7 and 8.

The first finding of the change records in Table 4.1 is that multiple changes that occur on the same day always can be corroborated with detected change-points. For example, in House 55 two changes in the sensor system were found to occur on 2018-08-08 when a change-point is detected also for this date (see Table 3.6, Chapter 3). Other such days with multiple changes and can be corroborated with detected change-points include 2017-04-03 for House 14 and 2018-10-24 for House 53.

To get an overview of the change records in Table 4.1, we summarize the number of change records in terms of maintenance type and the alias of the motion sensors respectively and the percentage of how many of them can be corroborated with the detected change-points in Table 4.2A. Let D_i be the set of all possible alias $i's$, i.e., 'In Living Room', 'In Front Door Area', 'In Bedroom, Master', 'In Bedroom, Master (General)', 'In Bathroom, Master', 'In Kitchen', 'In

Shower', 'In Dining Area', 'In Office Area', and 'In Bathroom, Guest', D_j be the set of all possible $j's$, i.e., 'Add a new sensor', 'Adjust the view', 'Reinstall', 'Relocate', 'Replace battery', and 'Replace with another sensor'. Let a_{ij} represent the number of maintenance type j (used as the columns in Table 4.2A) that occurred on the motion sensor with alias i (used as the rows in Table 4.2A). The value of a_{ij} is the number within the parenthesis at the intersection of row i and column j in Table 4.2A. For example, '(3)' at the intersection of row 2 and column 3 represents that there are 3 'Reinstall' records in maintenance log for the motion sensor with the alias 'In Front Door Area'. The number 2 above '(3)' represents that two of these three maintenances can be corroborated with the detected change-points.

We calculate the percentage of the record changes for a given sensor with alias i can be corroborated with the detected change-points. For example, there are four 'Add a new sensor' occurring on the motion sensor 'In Bedroom, Master (General)' with three of them can be corroborated with detected change-points; for the same sensor, there are two 'Replace battery' records in the maintenance log with both of them can be corroborated with detected change-points. No other types of maintenance log are found for this sensor. Thus the percentage is (3+2)/(4+2)*100% = 83.33%.

Similarly, we calculate the percentage of the record changes for a given maintenance log j can be corroborated with the detected change-points. For example, there are in total 1/3+1/3+1/3+1 = 2 records for 'Adjust the view' over all possible sensor alias (the summation of the numbers in the parenthesis in column 2 in Table 4.2A), among which 1/3+1/3+1/3 = 1 can be corroborated with a detected change point. Thus the percentage is 1/2*100% = 50%.

For multiple records on the same day, assuming that there are n records, we assign $1/n$ change to each record as the number of changes that record brings to the data collection environment. For example, in Table 4.1 we know that there are three ‘Adjust the view’ maintenance records for three motion sensors (‘In Front Door’, ‘In Bedroom, Master’, and ‘In Shower’) in House 014 on 2017-04-03. According to this rule, the number of changes that each motion sensor brings to the system is ‘1/3’ in column 2 in Table 4.2A.

In summary, 14 out of 42 change records (33.33%) were corroborated with detected change-points. In terms of the maintenance type, 50% of ‘Adjust the view’, 50% of ‘Reinstall’, and 63.64% of ‘Add a new sensor’ can be corroborated detected change-points; while for ‘Replacing batteries’, ‘Battery dead’, and ‘Replacing malfunctioned sensor’, the percentages are below 40%.

We further categorize the records in maintenance log and battery information in terms of their impact on the daily trajectories as shown in Table 4.2B. ‘Battery dead’ results in the absence of sensor events in the trajectory; ‘Add a new sensor’, ‘Reinstall’, ‘Replace battery’, and ‘Replace sensor’ result in the presence of new sensor events; while ‘Adjust the view’ and ‘Relocate sensor’ lead to changes in the observed events. Overall, 28.57% of records that indicate the absence of sensor events, 37.50% that indicate the presence of new sensor events, and 25% relates to the change in the observed events can be corroborated with detected change-points.

In terms of sensor alias, over 80% of the records on ‘In Bedroom, Master (General)’ (83.33%) and ‘In Bathroom, Master’ (100%) can be corroborated with detected change-points; while the percentage for sensors ‘In Living Room’, ‘In Front Door Area’, ‘In Bedroom, Master’, and ‘In Shower’ is only between 20% and 45%. None of the records for other sensors can be corroborated with the detected change-points. These results indicate that the detected change-points in the sequence of daily entropy rate shown in Chapter 3 cover partial of sensor system changes, which lead to the question of what changes in the sensor system are more likely to be detected by changes in the daily entropy rate.

In next section, we construct metrics to describe a change record based on the changes in ambient data before and after this change record; then use a classification decision tree to help identify the main metrics that distinguish the change records that can be corroborated with detected change-points from those that cannot.

4.3 Metrics Construction and Analysis for Change Records in Maintenance Logs

In this analysis, we included houses that have only one change record for a day and have at least one change record that can be corroborated with a detected change point. These houses include House 13, 27, 28, 51, and 54 and there are in total 24 change records, among which nine are the change records that can be corroborated with detected change-points, and 15 cannot.

We define some metrics (Table 4.3) to describe the collected ambient data before and after the change record. We first calculate an individual metric in Table 4.3 except $diff\_num$ for days before and after the change record respectively. Then use the difference of metrics for before and after the change record as a feature to describe the change record. For example, the calculated N_{min}'s for days before and after a change record is denoted as $N_{\min\_before}$ and $N_{\min\_after}$ respectively. The difference metric of N_{min} is calculated by $N_{\min\_diff} = N_{\min\_before} - N_{\min\_after}$. Similarly, we obtain other difference metrics including $N_{\max\_diff}$, N_{mode_diff}, $\bar{S}^{unc}{}_{diff}$, $\bar{S}^{real}{}_{diff}$, $\bar{l}$, $avg_{changed\_diff}$, and $prop_{changed\_diff}$. In addition to the above difference metrics, we also use is_top3_{before}, i.e., if the sensor events from the changed sensor rank top 3 before a change records, is_top3_{after}, i.e., if the sensor events of the changed sensor rank top 3 after a change record, as well as $diff\_num$, i.e., the count of different items between the set

$alias_{Top\ i\_before}$ and $alias_{Top\ i\_after}, i = 1, 2, 3$, to describe a change record. In total, there are eight difference metrics and three non-difference metrics to be used to describe a change record.

We use the eight difference metrics and is_top3_{before}, is_top3_{after}, and $diff\_num$ to train a classification decision tree [123] to classify the change records as $record_{cor}$ and $record_{non\_cor}$. The result shows that only $\bar{S}^{unc}{}_{diff}$ is used in the tree construction. The following if-then rules are generated by the decision tree model:

- If $\bar{S}^{unc}{}_{diff} \leq -0.32$, the change record is a $record_{cor}$.
- If $-0.32 < \bar{S}^{unc}{}_{diff} \leq 0.19$, the change record is a $record_{non\_cor}$.
- If $\bar{S}^{unc}{}_{diff} > 0.19$, the change record is a $recond_{cor}$.

The misclassification error for the training dataset is 8.3% and two $record_{non\_cor}$ are classified as $record_{cor}$. The confusion matrix is shown in Table 4.4.

These rules indicate that when the difference in the value of S^{unc} between the days before and after a change record is relatively large, the change records are more likely to be identified by the real entropy-based change-point algorithm. According to the definition, the value of S^{unc} is determined by the distribution of motion sensor events. In other words, when the changes in the sensor system result in significant changes in the probability distribution of the frequency of sensor events, these changes can be captured by the value of daily entropy rate. This is reasonable because changes in the distribution of sensor events illustrate possible changes in movement patterns and the frequency of these patterns which actually determine the value of the daily entropy rate.

Next, using House 27 as an example we visually examine the relationship between the changes in the distribution of sensor events and the detection of change-points in the sequence of daily entropy rate.

In House 27, there are nine change-points that divide the sequence of daily entropy rate into 10 segments as shown in Table 4.5. The bolded four dates in the column 'Date begin' are those that can be corroborated with detected change-points in the sequence of daily entropy rate. For each segment, we calculated the probability distribution of sensor events and graph them in Figure 4.1.

Let's first focus on the changes in the probability distributions of sensor events between four pairs of segments namely Segments 1 vs 2, Segments 3 vs 4, Segments 6 vs 7, and Segments 7 vs 8 whose transition points coincide with change-points detected by the algorithm proposed in Chapter 3.

In terms of Segment 1 vs Segment 2, from the battery information and maintenance log in Table 4.2A, we know that the battery of motion sensor with the alias 'In Living Room' was depleted and no sensor events were reported by this sensor during Segment 1. On 2017-01-24 when the battery was replaced the motion sensor reported data normally in Segment 2. In Figure 4.1 the probability of sensor events of the motion sensor with the alias 'In Living Room' is zero in Segment 1 and then rises to near 0.3 in Segment 2 after the dead battery is replaced. In addition, the probabilities of events for sensors 'In Bathroom, Master' and 'In Kitchen' are comparable in Segment 1 and both of them drop from about 0.4 to about 0.3 in Segment 2 due to the increased number of sensor events of the sensor 'In Living Room'.

In terms of Segment 3 vs Segment 4, the probabilities of sensor events of the motion sensor 'In Bedroom, Master (General)' and of the sensor 'In Shower' are 0.1 and 0.25 respectively in Segment 4. However, both of these are almost zero in Segment 3. These changes in the probability distribution are introduced by the installation of these two sensors in Segment 4.

In terms of Segment 6 vs Segment 7 and Segment 7 vs Segment 8, both probabilities of sensor events of the sensor 'In Living Room' and the sensor 'In Bedroom, Master (General)' decrease to almost zero in Segment 7 from 0.3 and 0.15 respectively in Segment 6, and then in Segment 8, they rise back to 0.25 and 0.15. These changes in the probability distributions of sensor events correspond to the records in the maintenance log and battery information that the batteries are dead and are then replaced for the sensor 'In Living Room' and the sensor 'In Bedroom, Master (General)'.

We note that all the probability distribution changes between four pairs of segments introduced above that can be corroborated with detected change-points relate to the presence or absence of sensor events of motion sensors that are installed in the living room and the master bedroom where the independent resident normally spends considerable time during his/her daily life. However, the presence of new sensor events or the absence of sensor events is not always detected by the algorithm proposed in Chapter 3. For example, in Segment 4 the probability of sensor events for the sensor 'In Shower' is 0.25 while in Segment 5 the probability is near zero and this change does not coincide with any of the change-points discovered in Chapter 3, implying that this probability change relating to the sensor 'In Shower' did not cause a significant change in the value of daily entropy rate.

In addition to visually examining the probability distributions, we can also quantitatively measure the extent to which the two probability distributions of sensor events in two adjacent segments are different using distance measures of histograms such as χ^2 statistic [124].

Given two histograms $H = \{h_i\}$ and $K = \{k_i\}$, χ^2 statistic can be calculated using

$$d_{\chi^2}(H, K) = \sum_i \frac{(h_i - m_i)^2}{m_i}$$

where $m_i = \frac{h_i - k_i}{2}$.

Table 4.6 shows the calculated dissimilarity measures between probability distributions of sensor events in each pair of two adjacent segments. For χ^2 statistic, all four pairs of adjacent segments (bolded in Table 4.6) that relate to change records that were discovered by detected change-points have a dissimilarity measure bigger than 0.2. In other words, when two probability distributions before and after a change record are sufficiently dissimilar, the change record can be detected by changes in the daily entropy rate.

4.4 Overview of Presence of Visitors in Bi-weekly Assessments

In this section, we focus on the records in bi-weekly assessments related to long-term visitors. In most cases, this information pertains only to visitors who stay with the participant multiple days/weeks, and in many cases the start and end dates of the visit are approximations. Visitor information is summarized in Table 4.7. Notice that either arriving or leaving of visitors counts for a change in the data collection environment. In total, we determine 13 visitor arrival and departure events with the start and end date recorded and 7 of them coincide with detected change-points.

4.5 Metrics Construction and Analysis for Presence of Visitors in Bi-weekly Assessments

Similar to Section 4.3, we construct metrics in this section to help us identify the key features that distinguish visitor events that were discovered with detected change-points from those that were not. The metrics we use in this section are listed in Table 4.8. In addition to the six difference metrics including $N_{\min\_diff}$, $N_{\max\_diff}$, N_{mode_diff}, $\bar{S}^{unc}{}_{diff}$, $\bar{S}^{real}{}_{diff}$, $\bar{l}$, and $diff\_num$ that we use in Section 4.3, we introduce two new metrics for visitor events in this section: Num_{alias} defined as the mean of daily sensor events for motion sensor with a specific alias and $Prop_{\text{alias}}$ defined as the mean of the daily proportion of sensor events for a motion sensor with a specific alias. For these two metrics, instead of using difference of metrics for days before and after a change record to describe a change record, we use the absolute value of the difference metric to capture the absolute changes that visitors' arrival and departure introduce to the ambient sensor data. To be specific, metrics include $|Num_{\mathrm{alias\_diff}}|$ and $|Prop_{\mathrm{alias\_diff}}|$, $alias$ = 'In Living Room', 'In Kitchen', 'In Bedroom, Guest', 'In Bathroom, Guest', 'In Front Door Area', 'In Bedroom, Master', or 'In Bathroom, Master' are used for analysis. We do not use is_top3_{before} and is_top3_{after} in this analysis because the changes are not related to a specific sensor but the overall system.

We use t-test to compare the mean of difference metrics for two groups, i.e., $record_{cor}$ and $record_{non\_cor}$. Results show that the mean of daily sensor events and the mean of the proportion of daily sensor events for motion sensors 'In Living Room' and 'In Bathroom, Guest' are significantly different between two groups (Table 4.9). Both the daily sensor events in the living room and the guest bathroom for $record_{cor}$ are almost doubt of those for $record_{non\_cor}$, illustrating that visitors' activities bring significant changes to the number and proportion of sensor events in the living room and the guest bathroom which can be captured by the changes in the value of daily entropy rate. Due to the increase in the number and proportion of sensor events in the living room and guest bathroom the proportions of the sensor events of motion sensors 'In Bedroom, Master' and 'In Front Door Area' are significantly lower than those when the visitors are absent.

Similarly, as in Section 4.3, we use a classification decision tree to help determine the feature that can distinguish visitor records that can be corroborated detected change-points from those that cannot. We train a classification decision tree use the features we constructed for each visitor event, and the result shows that only $|Num_{\text{Bathroom,Guest\_diff}}|$ is used for tree construction. The tree branches can be interpreted as the below rule:

- $\left|Num_{\text{In Bathroom,Guest}_{\text{diff}}}\right| > 14$, the visitor record is a $record_{cor}$; otherwise, the record is a $record_{non\_cor}$.

The misclassification error for the training dataset is 0, meaning that all $record_{cor}$ have the value of $\left|Num_{\text{In Bathroom,Guest}_{\text{diff}}}\right|$ bigger than 14. It could be possible that when the number of sensor events in the guest bathroom exceeds some threshold, the new mobility patterns that are brought by the increased number in the sensor events of the motion sensor installed guest bathroom begins to count in the changes in the value of daily entropy rate.

4.6 Conclusion

In this chapter, we analyze the records about system changes in the maintenance log and the presence of visitors in the bi-weekly assessments. We observe that 33.33% of records in the maintenance log and 53.85% records in the bi-weekly assessment about the presence of visitors are detected by the detected change-points in the sequence of daily entropy rate described in Table 3.6 in Chapter 3.

We construct metrics to describe records about the sensor system changes and the presence of visitors respectively. We compare the records that coincide with detected change-points in daily entropy rate and those that do not and investigate the distinguishing characteristics between these groups. The rules generated by a classification decision tree on a dataset of 24 change records indicate that the change record is more likely to be detected when the difference in the value of daily temporal uncorrelated entropy rate between two adjacent segments exceeds a threshold, indicating change records that bring significant changes in probability distribution of sensor events are more likely to be detected by changes in daily entropy rate. These changes could be caused by the presence or absence of sensor events of motion sensors installed in areas where the resident is more likely to spent considerable time during his/her daily life such as the living room or the master bedroom are more likely to be detected (Table 4.2A), and the changes in the presence of sensor events that relate to activities such as taking a shower or cooking that occur less frequently during daily life are less likely to be detected by daily entropy rate. In the view of system maintenance, this finding suggests new methods and metrics in addition to daily entropy rate should be considered to effectively identify the undetected changes, such as threshold method based on the dissimilarity measures of probability distributions. This is important especially for the changes from sensors that report a small number of sensor events in normal days (the sensor 'In Shower') but could closely relate to a person's health status.

For the visitor records, the detected records that can be corroborated with the detected change-points in the sequence of daily entropy rate are those that have significant differences in the numbers and proportions of sensor events in the living room, the guest bathroom, the front door area, and the master bedroom between days before and after the visitors' arrival and departure. The rules obtained from a classification tree indicate that the daily movement level (the number of daily sensor events) in the guest bathroom is much higher for records that can be corroborated with detected change-points than those that cannot. In addition, we notice that all the detected visitor records relate to either multiple family members or a long-term staying of a family member; and four out of six undetected visitor records relate to friends and their short-term stay. This suggests that the length of the visit period and the number of visitors bring different levels of changes in the collected ambient sensor data. Collecting and analyzing data of different types of visitors from various houses could help us design algorithms that can effectively detect the presence and absence of visitors in real living environments.

Table 4.1
Changes in sensor systems and device battery information over two years (2017 and 2018)

House	Date	Alias of motion sensors (illustrate the location and function of sensors	Maintenance type	Description	Is the maintenance corroborated with a detected change-point?
	2017-01-24	In Bathroom, Guest	Relocate	The location of this sensor might be too high for the participant. Changed its location next to the door and lowered it down.	No
	2017-05-17	In Living Room	Replace with another sensor	The old sensor was 'stuck on' at random times.	No 1/2
	2017-05-17	In Shower	Replace battery	The battery run out and there was no data since 2017-02-14.	No 1/2
	2017-02-14	In Shower	Battery dead	No data reported due to the battery running out.	No
	2017-08-15	In Bedroom, Master (General)	Add a new sensor	Add a new sensor for a better coverage	No
	2017-10-24	In Front Door Area	Reinstall	The magnet on the front door was kicked off. It was re-installed on 2017-10-24.	No
	2018-09-26	In Front Door Area	Replace battery	The battery run out and there was no data since 2019-09-13.	No
8	2018-09-13	In Front Door Area	Battery dead	No data reported due to the battery running out	No
	2017-01-24	In Living Room	Replace battery	The battery run out and there was no data since 2017-01-06	No
	2017-01-03	In Living Room	Battery dead	No data reported due to the battery running out	No
	2017-08-16	In Bedroom, Master (General)	Add a new sensor	Add a new sensor for a better coverage	Yes
	2018-06-11	In Office Area	Replace battery	The battery run out and there was no data since 2018-06-07	No
13	2018-06-07	In Office Area	Battery dead	No data reported due to the battery running out.	No
	2017-04-03	In Front Door Area	Adjust the view	Lower it down and repaired it again.	Yes 1/3
	2017-04-03	In Bedroom, Master	Adjust the view	Lower it down to better capture the participant's motion	Yes 1/3
14	2017-04-03	In Shower	Adjust the view	Lower it down to better capture the participant's motion.	Yes 1/3

Table 4.1 (Continued)

	2017-01-24	In Living Room	Replace battery	The battery run out and there was no data since 2016-12-20.	Yes
	2017-04-12	In Bathroom, Guest	Adjust the view	Lowered it down to better capture motion.	No
	2017-08-15	In Bedroom, Master (General)	Add a new sensor	Add a new sensor for a better coverage	Yes
	2017-09-26	In Shower	Replace battery	The battery run out since 2017-08-19.	No
	2017-08-19	In Shower	Battery dead	No data reported due to the battery running out.	No
	2018-03-13	In Bedroom, Master (General)	Replace battery	The battery run out and there was no data since 2018-01-14.	Yes
	2017-12-22	In Bedroom, Master (General)	Battery dead	No data reported due to the battery running out.	Yes
	2018-09-14	In Living Room	Battery dead	The battery run out and there was no data between 2018-09-14 and 2019-03-05	No
27	2018-11-18	In Shower	Battery dead	The battery run out and there was no data since 2018-11-18.	No
	2018-08-15	In Bedroom, Master (General)	Add a new sensor	Add a new sensor for a better coverage	Yes
	2018-02-22	In Kitchen	Relocate	The device has been re-positioned due to its coverage area. Less overlapping with the living room.	No
	2018-09-05	In Shower	Replace battery	The battery run out and there was no data since 2018-05-12.	No
28	2018-05-12	In Shower	Battery dead	No data reported due to the battery running out.	No
	2018-06-11	In Living Room	Replace with another sensor	USF.AL.MS.184 was replaced with AN.AL.MS.78 because this sensor stuck on sometimes.	Yes
	2018-09-26	In Shower	Replace battery	The battery run out and there was no data since 2018-08-27.	No
51	2018-08-27	In Shower	Battery dead	No data reported due to the battery running out.	No
53	2018-06-11	In Kitchen	Replace with another sensor	USF.EL.MS.443 was replaced with USF.EL.MS.563 because this sensor stuck on most of the time.	No

Table 4.1 (Continued)

	2018-10-24	In Living Room	Replace battery	The battery run out and there was no data since 2018-10-05.	Yes 1/2
	2018-10-24	In Shower	Replace battery	The battery run out and there was no data since 2018-09-28.	Yes 1/2
	2018-10-05	In Living Room	Battery dead	No data reported due to the battery running out.	No
53	2018-09-26	In Shower	Battery dead	No data reported due to the battery running out.	Yes
	2018-08-09	In Dining Area	Add a new sensor	Add a new sensor to get better coverage	No
	2018-09-26	In Front Door Area	Reinstall	The device was kicked off and no data since 2018-9-12.	Yes
54	2018-09-12	In Front Door Area	Battery dead	No data reported by the motion sensor cause it was kicked off.	Yes
	2018-08-08	In Bathroom, Master	Add a new sensor	Add a new sensor for better coverage	Yes 1/2
	2018-08-08	In Living Room	Replace with another sensor	USF.AL.MS.630 was replaced with USF.AL.MS.632 because this device cannot be paired/repaired. No data was reported since 2018-07-19.	Yes 1/2
	2018-07-19	In Living Room	Battery dead	No data reported by the motion sensor 'In living room'.	No
55	2018-10-02	In Shower	Replace with another sensor	AN.AL.MS.77 was replaced with USF.AL.MS.513 because it didn't report motion.	No
	2018-10-08	In Living Room	Replace battery	No data after 2018-10-08 and this sensor was repaired on 2019-01-22.	No 1/3
	2018-10-08	In Kitchen	Replace battery	The battery run out and there was no data since 2018-09-30.	No 1/3
	2018-10-08	In Shower	Replace battery	The battery run out and there was no data since 2018-09-08.	No 1/3
	2018-09-30	In Kitchen	Battery dead	No data reported by the motion sensor 'In kitchen'.	No
56	2018-09-08	In Shower	Battery dead	No data reported by the motion sensor 'In shower'.	Yes

Table 4.2A
A detailed summary of change records in terms of the maintenance type and alias of the motion sensors

Alias	Battery dead	Add a new sensor	Reinstall	Replace battery	Replace sensor	Adjust the view	Relocate	Percentage that sensor with alias *i* corroborated with change-points
In Living Room	0 (3)	-	-	1+1/2 (3+1/2+ 1/3)	1+1/2 (2)	-	-	(3)/(8+1/2+1/3)= 33.96%
In Front Door Area	1 (2)	-	1 (2)	0 (1)	-	1/3 (1/3)	-	(2+1/3)/(5+1/3)= 43.75%
In Bedroom, Master	-	-	-	-	-	1/3 (1/3)	-	(1/3)/(1/3)= 33.33%
In Bedroom, Master (General)	1 (1)	3 (4)	-	1 (1)	-	-	-	(3+2)/(4+2)= 83.33%
In Bathroom, Master	-	1/2 (1/2)	-	-	-	-	-	(1/2)/(1/2)= 100%
In Kitchen	0 (1)	-	-	0 (1/3)	0 (1)	-	0 (1)	(0)/(3+1/3)=0%
In Shower	2 (6)	-	-	1/2 (5+1/3)	0 (1)	1/3 (1/3)	-	(2+1/2+1/3)/(12+2/3)= 22.37%
In Dining Area	-	0 (1)	-	-	-	-	-	(0)/(1)= 0%
In Office Area	0 (1)	-	-	0 (1)	-	-	-	(0)/(2)= 0%
In Bathroom, Guest	-	-	-	-	-	0 (1)	0 (1)	(0)/(2)= 0%
Percentage that maintenance type (*j*) corroborated with change-points	(4)/(14) = 28.57%	(3+1/2)/ (5+1/2) = 63.64%	(1)/(2) = 50%	(3)/(12+ 1/2) = 24%	(1+1/2)/(4) = 37.50%	(1)/(2) = 50%	0/2 = 0	(14)/42 = 33.33%

Table 4.2B
Aggregate summary of change records in terms of their impact of daily trajectories

Type of impact on data collection	Absence of sensor events	Presence of new sensor events	Change in the observed events
Percentage that impacts type (*k*) corroborated with change-points	(4)/(13)=28.57%	(9)/(24) = 37.50%	(1)/(4) = 25%

Table 4.3
Metrics that are constructed for days before and after a record about sensor system changes

Metrics	Value Type	Description
N_{min}	Integer	The minimum number of distinct motion sensors that fire on a day
N_{max}	Integer	The maximum number of distinct motion sensors that fire on a day
N_{mode}	Integer	The mode of the number of distinct motion sensors that fire on a day
$\bar{S}^{unc}$	Numerical	The mean of temporal-uncorrelated entropy rate of daily trajectories
$\bar{S}^{real}$	Numerical	The mean of real entropy rate of daily trajectories
$\bar{l}$	Numerical	The mean of the length of daily trajectories
$alias_{Top\ i}, i = 1,2,3$	Categorical	The alias of the sensor with the number of sensor events ranking top i including 'In Living Room', 'In Bedroom, Master', 'In Bathroom, Master', 'In Bedroom, Guest', 'In Bathroom, Guest', etc.
$avg_{changed}$	Integer	The daily average number of sensor events of the changed sensor
$prop_{changed}$	Numerical	The daily average proportion of sensor events of the changed sensor
is_top3_{before}	Binary	$is\_top3_{before} = 1$ represents that the number of sensor events for the changed sensor ranks among top 3 BEFORE a change record; $is\_top3_{before} = 0$ represents that the number of sensor events for the changed sensor doesn't rank among the top 3 BEFORE a change record.
is_top3_{after}	Binary	$is\_top3_{after} = 1$ represents that the number of sensor events for the changed sensor ranks among top 3 AFTER a change record; $is\_top3_{after} = 1$ represents that the number of sensor events for the changed sensor doesn't rank among the top 3 AFTER a change record.
$diff\_num$	Integer	$diff\_num$ = 0, 1, 2, 3 represents the count of different items between the set $alias_{Top\ i\_before}$ and $alias_{Top\ i\_after}, i = 1, 2, 3$.

Table 4.4
Confusion matrix of the classification decision tree

	Actual $record_{cor}$	Actual $record_{non\_cor}$
Predicted $record_{cor}$	9	2
Predicted $record_{non\_cor}$	0	13

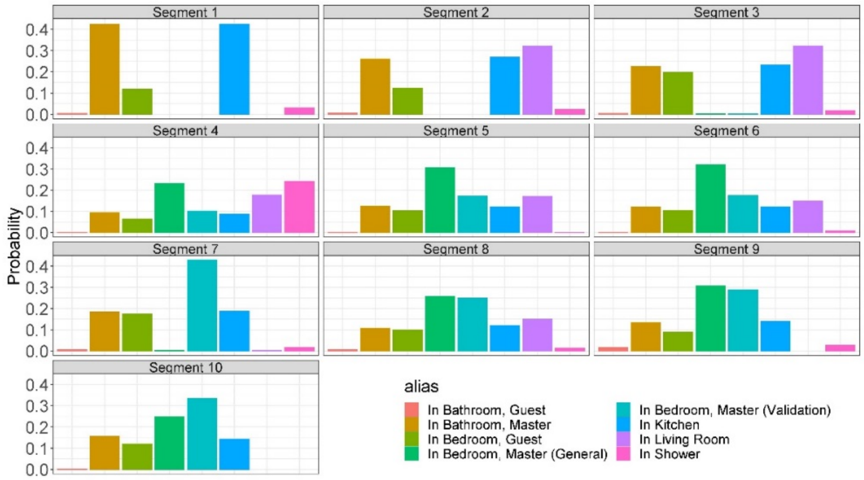


Figure 4.1 Probability distributions of motion sensor events in different segments (House 27).

Table 4.5
Ten segments determined by nine change records in the maintenance log and battery information (House 27). Bolded dates are those can be corroborated with detected change-points in daily entropy rate

Segment	Date begin	Date end	Days	Battery info and maintenance Alias	Impact on data collection
1	2017-01-01	2017-01-24	21	Battery was dead In Living Room	Absence of events
2	**2017-01-25**	2017-04-12	78	Replace battery In Living Room	Presence of new events
3	2017-04-13	2018-08-15	115	Adjust the view In Bathroom, Guest	Change of observed events
4	**2017-08-16**	2017-08-19	3	Add a new sensor In Bedroom, Master (General)	Presence of new events
5	2017-08-20	2017-09-26	38	Battery dead In Shower	Absence of events
6	2017-09-27	2017-12-22	85	Replace battery In Shower	Presence of new events
7	**2017-12-23**	2018-03-13	57	Battery dead In Bedroom, Master (General)	Absence of events
8	**2018-03-14**	2018-09-14	176	Replace battery In Bedroom, Master (General)	Presence of new events
9	2018-09-15	2018-11-18	63	Battery dead In Living Room	Absence of events
10	2018-11-19	2018-12-31	38	Battery dead In Shower	Absence of events

Table 4.6
The distance of the probability distributions between two adjacent segments (House 27)

The segment before the change record vs. the segment after the change record	χ^2 statistic
Segment 1 vs. Segment 2	**0.20**
Segment 2 vs. Segment 3	0.013
Segment 3 vs. Segment 4	**0.37**
Segment 4 vs. Segment 5	0.14
Segment 5 vs. Segment 6	0.0049
Segment 6 vs. Segment 7	**0.31**
Segment 7 vs. Segment 8	**0.25**
Segment 8 vs. Segment 9	0.086
Segment 9 vs. Segment 10	0.028

Table 4.7
The presence of visitors over two years (2017 and 2018) according to the bi-weekly assessment

House	Visitors arrive	Visitors leave	Description	Arriving date corroborated with detected change-points	Leaving date corroborated with detected change-points
8	2017-01-25	2017-02-28	Son has been visiting, granddaughter and great-granddaughter came for four days in mid-February	Yes	Yes
8	2017-08-07	2017-08-10	Friend from out of town was visiting	No	No
8	2017-09-09	2017-09-12	Friends stayed during the storm	No	No
8	2017-11-03	2017-11-13	Son and daughter are visiting for an unspecified amount of time. Granddaughter and her boyfriend are also coming when the parents are there.	Yes	Yes
8	2018-12-18	2018-12-25	Grandson visited.	No	No
13	2018-02-11	2018-02-24	Daughter visited	Yes	Yes
55	2018-11-22	NA	Visitors may still be present after the last day included in the analysis	Yes	-

Table 4.8
Metrics that are constructed for the days before and after a change record about visitors

Metrics	Value Type	Description
N_{min}	Integer	The minimum number of motion sensors that fire on a day
N_{max}	Integer	The maximum number of distinct motion sensors that fire on a day
N_{mode}	Integer	The mode of the number of distinct motion sensors that fire on a day
$\bar{S}^{unc}$	Numerical	The mean of temporal-uncorrelated entropy rate of daily trajectories
$\bar{S}^{real}$	Numerical	The mean of real entropy rate of daily trajectories
$\bar{l}$	Numerical	The mean of the length of daily trajectories
$alias_{Top\ i}, i = 1,2,3$	Categorical	The alias of the sensor with the number of sensor events ranking top i including 'In Living Room', 'In Bedroom, Master', 'In Bathroom, Master', 'In Bedroom, Guest', 'In Bathroom, Guest', etc.
$diff\_num$	Integer	$diff\_num$ = 0, 1, 2, 3 represents the count of different items between the set $alias_{i\_before}$ and $alias_{i\_after}, i = 1, 2, 3$.
Num_{alias}	Integer	The daily mean of sensor events of the motion sensor with $alias_i$, $alias_i$ = 'In Living Room', 'In Kitchen', 'In Bedroom, Guest', 'In Bathroom, Guest', 'In Front Door Area', 'In Bedroom, Master', or 'In Bathroom, Master'
$Prop_{alias}$	Numerical	The daily proportion of sensor events of the motion sensor with $alias_i$, $alias_i$ = 'In Living Room', 'In Kitchen', 'In Bedroom, Guest', 'In Bathroom, Guest', 'In Front Door Area', 'In Bedroom, Master', or 'In Bathroom, Master'

Table 4.9
Mean and standard deviation (SD) of difference metrics for $record_{cor}$ and $record_{non\_cor}$ in bi-weekly assessments about visitors

Difference metrics	Mean (SD) for $record_{cor}$ vs. Mean (SD) for $record_{non\_cor}$	P-value
$\lvert Num_{\text{In Living Room}\_diff}\rvert$	12.57 (7.4) vs. 29 (15.42)	0.049
$\lvert Prop_{\text{In Living Room}\_diff}\rvert$	0.89% (0.57%) vs. 7.67% (4.88%)	0.019
$\lvert Num_{\text{In Bathroom,Guest}\_diff}\rvert$	19.43 (3.55) vs. 11 (2.53)	4.54e-4
$\lvert Prop_{\text{In Bathroom,Guest}\_diff}\rvert$	0.59% (0.38%) vs. 2.59% (1.73%)	0.036
$\lvert Prop_{\text{In Bedroom,Master}\_diff}\rvert$	0.89% (0.63%) vs. 4.48% (1.56%)	1.51e-3
$\lvert Prop_{\text{In Front Door Area}\_diff}\rvert$	1.93% (1.68%) vs. 12.44% (7.47%)	0.017

Chapter 5: Future Work

In this section, we introduce future research directions

5.1 Available Features and Methods Used for System Change Detection

Changes and disruptions in the data collection environments are unavoidable when using ambient sensor networks for activity monitoring in the real living environments, introducing deviations in the collected ambient sensor data and misunderstanding of the occupants' daily activities. Our results in Chapter 4 indicate that changes in the daily real entropy rate detected by minimizing a penalized contrast function are able to identify part of the changes and disruptions in the data collection environments. To discover the undetected system changes and facilitate the long-term reliability of sensor systems for data collection and health monitoring, other features of sensor data and change-point detection algorithms could be considered for the detection of system changes.

System changes such as sensors' failing to report data change the number of sensor events and therefore their probability distribution. Quantitative metrics that describe the probability distributions of sensor events may be useful features to capture such changes in the system. Recall that the entropy is a measurement that quantifies the uncertainty or information in a random variable. Defining the resident's location as a random variable, the entropy estimated from the distribution of sensor events quantitatively measures the uncertainty of the resident's location which is represented by sensor events. When the resident spends his/her majority time in the living room and much less time in other locations, the probability for the sensor events relating to the living room will be the biggest in the distribution and the uncertainty of the resident's location measured by the entropy will be small. When the sensors in the living room malfunction, the probability distributions constructed by other sensor events will become uniform and thus the uncertainty of the resident's location measured by the entropy will increase. The changes in the entropy measurement illustrate the status changes of the sensors installed in the living room and could be used as features to track system changes.

It is illustrated in our previous study (Chapter 4) that the uncorrelated entropy rate is the dominant feature to distinguish the detected changes from the undetected ones, suggesting that the uncorrelated entropy rate could be a useful feature to track system changes. Besides the information measurements estimated from a single probability distribution such as the uncorrelated entropy rate and the real entropy rate, features that measure the difference in the information measurements of two probability distributions, i.e., the information gain, or the dissimilarity of two probability distributions such as χ^2 statistic and the relative entropy or Kullback-Leibler distance [117] can be useful to quantitatively measure changes that occur in distributions. Other available features include the shape characteristics of probability distribution such as skewness and kurtosis. These features capture different aspects of the collected data and we could compare their performance on the detection of change-points and identify features that are useful to detect changes that are not discovered by merely using the real daily entropy rate.

In the study of human mobility, we modeled the sequential data of daily real entropy as piecewise constant and estimated the unknown abrupt changes in the mean by minimizing a penalized contrast function. In the current literature, many other alternative methods are available for the detection of unknown changes based on the statistical features of the data. A classical idea is to compare the probability distributions of data before and after the candidacy change-points based on hypothesis testing. A null hypothesis that there is no change occurring in the sequential data and an alternative hypothesis that the probability distribution changes at some points are the typical hypotheses for this method. Test statistics, for instance, the likelihood ratio, i.e., the ratio of likelihood functions [125] is estimated based on observations and the assumed probability distribution to help determine the change points [126]. Other change-point detection methods can be found in [126]–[128]. These methods provide us alternative models to fit the sequential data for the discovery of undetected system changes. A comparison of how these methods perform on the detection of system-changes could be one of the future research directions.

5.2 Real-time Location and Activity Tracking

In smart homes, real-time analysis and tracking of smart home occupant's location and activity are necessary to provide services to the occupants' needs. One of our future work is to transfer the current offline analysis algorithms to real-time algorithms to facilitate real-time smart home applications.

In our study of using ambient sensor networks to monitor the occupant's health and wellness outlined in Chapter 2, an offline algorithm is proposed to track the occupant's location and status and facilitate a retrospective analysis of the occupant's activity routines and health conditions.

Based on this offline algorithm, we designed a rule-based online algorithm to track the occupant's location and outings in real time using ambient sensor data. In this algorithm, shown in Figure 5.1, the occupant's locations are maintained in a location ordered list which is updated based on rules and the sensor events reported by motion sensors and contact sensors that are installed on the exit doors. The algorithm in Figure 5.1 tracks the resident's locations and his/her outing activities in real time. In the future, real-time tracking of other health-related activities such as sleep behaviors will be incorporated.

5.3 Using Ambient Sensor Data for Health Monitoring

The location and status tracking algorithm proposed in this dissertation enables us to generate health-related indicators from ambient sensor data. By visualizing these indicators, we can explore an individual's life routines as well as his/her health conditions. However, visually examining graphs to determine health conditions is subjective and time and labor-consuming, and not scalable. In our future work, we would like to focus on quantitative relationships between the indicators and health conditions to facilitate the automatic monitoring and decision making about an individual's health condition. Some of our work may include the exploration of how indicators change related to changes in health conditions (if a higher duration of outings relates to a lower risk of developing depression and loneliness, if longer sleep durations and lower number of sleep disruptions relate to better sleep quality; if longer durations of outings relates to lower risk of loneliness) and how a quantitative evaluation of an individual's physical or mental health condition can be predicted by ambient indicators (if a score that evaluates sleep quality can be predicted by predictors such as sleep duration, times of sleep disruption, and time spent out of home). Further, in the study of human mobility, we discovered that the regularity of mobility, i.e., the daily real entropy rate, is positively correlated with age, i.e., the older age cohorts have a higher value of entropy rate. Considering the possible decline in physical and mental functionality due to aging, we could consider the entropy rate as an indicator and explore its relationship with an individual's physical and mental health condition as aging in future work.

Step 1: Create a blank Location Ordered List (LOL)
Step 2: Iterate through the list of events for all motion sensors and home exit sensors (contact sensors on the garage, front door, etc.) for a given period starting at the beginning and update the LOL according to rules as below:

    Rule 1: When there is a motion sensor ON event, add the motion sensor to the last position of LOL and remove any home exit events (if applicable)
    Rule 2: When there is a motion sensor OFF event, remove the motion sensor from the list (if applicable)
    Rule 3: When there is a home exit sensor ON event, add an exit sensor to the first position in the list if the first position is not already Exit.

Step 3: The last location in LOL is taken as the estimation of the occupant's current location.

Figure 5.1 A real-time location and activity tracking algorithm.

References

[1] A. W. Roberts, S. U. Ogunwole, L. Blakeslee, and M. A. Rabe, "The population 65 years and older in the United States: 2016," U.S. Department of Commerce Economics and Statistics Administration, U. S. Census Bureau, Washington, D.C., 2018. [Oneline]. Available: https://www.census.gov/content/dam/Census/library/publications/2018/acs/ACS-38.pdf

[2] "Global strategy and action plan on ageing and health," World Health Organization, Geneva, Switzerland, 2017. [Online]. Available: https://www.who.int/ageing/WHO-GSAP-2017.pdf?ua=1

[3] "World population ageing 2015," Department of Economic and Social Affairs, Population division, United Nations, New York, 2015. [Online]. Available: https://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2015_Report.pdf

[4] W. S. Pearson, K. Bhat-Schelbert, and J. C. Probst, "Multiple chronic conditions and the aging of America: Challenge for primary care physicians," *J. Prim. Care Community Health*, vol 3, pp. 51–56, Sept. 2011.

[5] "The state of aging & health in America 2013," Centers for Disease Control and Prevention, US Department of Health and Human Services, Atlanta, GA, 2013. [Online]. Available: https://www.cdc.gov/aging/pdf/state-aging-health-in-america-2013.pdf

[6] Health affairs: Health spending projections through 2027, Health Affairs Blog, Feb. 22, 2019. [Online]. Available: https://www.healthaffairs.org/do/10.1377/hblog20190221.997607/full/

[7] C. Hoffman, D. Rice, and H. Sung, "Persons with chronic conditions: Their prevalence and costs," *JAMA*, vol. 276, no. 18, pp. 1473-1479, Nov. 2012.

[8] M. Porter, "What is value in health care?," *N. Engl. J. Med.*, vol. 363, no. 26, pp. 2477-2481, Dec. 2010.

[9] M. Mather, L. A. Jacobsen, and K. M. Pollard, "Aging in the United States," *Popul. Bull.*, vol. 70, no. 2, pp. 2-17, Dec. 2015. [Online]. Available: https://www.prb.org/wp-content/uploads/2016/01/aging-us-population-bulletin-1.pdf

[10] Healthy Places Terminology, Centers for Disease Control and Prevention, 2017. [Online]. Available: https://www.cdc.gov/healthyplaces/terminology.htm

[11] E. Graybill, P. McMeekin, and J. Wildman, “Can aging in place be cost effective? A systematic review,” *PlOS ONE* 9(7):e102705, July 2014.

[12] J. L. Wiles, A. Leibing, N. Guberman, J. Reeve, and R. E. S. Allen, “The meaning of “aging in place” to older people,” *The Gerontologist*, vol. 52, no. 3, pp. 357–366, June 2012.

[13] K. Grimmer, D. Kay, J. Foot, and K. Pastakia, “Consumer views about aging in place,” *Clin. Interv. Aging*, vol. 10, pp. 1803-1811, Nov. 2015.

[14] L. F. Carver, R. Beamish, S. P. Phillips, and M. Villeneuve, “A scoping review: social participation as a cornerstone of successful aging in place among rural older adults,” *Geriatrics (Basel)*, vol. 3, no. 75, Dec. 2018.

[15] D. Kaplan, T. Andersen, A. Lehning, and T. E. Perry, “Aging in place vs. relocation for older adults with a neurocognitive disorder: applications of wiseman’s behavioral model,” *J. Gerontol. Soc. Work*, vol. 58, no. 5, pp. 521–538, July 2015.

[16] W. Mills, T. Regev, M. Kunik, N. Wilson, J. Moye, L. McCullough, and A. Naik, “Making and executing decisions for safe and independent living (MED-SAIL): development and validation of a brief screening tool,” *Am. J. Geriat. Phychiat.*, vol. 22, no. 3, pp. 285–293, Mar. 2014

[17] J. Augusto, V. Callaghan, D. Cook, A. Karneas, and I. Satoh, “Intelligent environments: A manifesto,” *Hum-cent. Comput. Info.*, vol. 3, no. 12, pp. 1-18, July 2013.

[18] O. Anya and H. Tawfik, “Leveraging big data analytics for personalized elderly care: opportunities and challenges,” in *Applied Computing and in Medicine and Health*, Elsevier, 2016, ch. 5, pp. 99-124.

[19] D. W. Bates, S. Saria, L. Ohno-Machado, A. Shah, and G. Escobar, “Big data in health care: using analytics to identify and manage high-risk and high-cost patients,” *Predictive Analytics*, vol. 33, no. 7, pp. 1123-1131, July 2014.

[20] M. Z. Uddin, W. Khaksar, and J. Torrensen, “Ambient sensors for elderly care and independent living: a survey,” *Sensors*, vol. 18, pp. 2027-2058, July 2018.

[21] A. Lotfi, C. Langensiepen, S. M. Mahmound, and M. J. Akhlaghinia, “Smart homes for the elderly dementia suffers: identification and prediction of abnormal behavior,” *J. Amb. Intel. Hum. Comp.*, vol. 3, no. 3, pp. 205-218, Jan. 2012.

[22] S. Majumder, E. Aghayi, M. Noferesti, H. Memarzadeh-Tehran, T. Mondal, Z. Pang, and M. J. Deen, “Smart homes for elderly healthcare-recent advances and research challenges,” *Sensors*, vol. 17, no. 11, Oct. 2017.

[23] D. Riboni, C. Bettini, G. Civitarese, Z. Janjua, and R. Helaoui, "SmartFABER: recognizing fine-grained abnormal behaviors for early detection of mild cognitive impairment," *Artif. Intell. Med.*, vol. 67, pp. 57-74, Feb. 2016.

[24] V. G. Sanchez, C. F. Pfeiffer, and N. Skeie, "A review of smart house analysis methods for assisting older people living alone," *J. Sens. Actuator. Netw.*, vol. 6, no. 11, pp. 1-38, July 2017.

[25] P. Urwyler, R. Stucki, L. Rampa, R. Muri, U. P. Mosimann, and T. Nef, "Cognitive impairment categorized in community-dwelling older adults with and without dementia using in-home sensors that recognize activities of daily living," *Sci. Rep.*, vol. 7, pp. 1-9, Feb. 2017.

[26] S. P. Rao, and D. J. Cook, "Prediction inhabitant action using action and task models with application to smart homes," *Int. J. Artif. Intell. Tools*, vol. 13, no. 1, pp. 81-99, 2004.

[27] G. Abowd, A. Bobick, I. Essa, E. D. Mynatt, and W. A. Rogers, "The aware home: A living laboratory for technologies for successful aging", In *Proc. AAAI-02 Workshop "Automation as Caregiver"*. July 2002.

[28] S. Helal, W. Mann H. El-Zabadani, J. King, Y. Kaddoura, and E. Jansen, "The gator tech smart house: A programmable pervasive space," *Comput.*, vol. 38, pp. 50–60, Mar. 2005.

[29] N. Agoulmine, M. J. Deen, J. Lee, and M. Meyyappan, "U-Health smart home," *IEEE Nanotechnol. Mag.*, vol. 5, no. 3, pp. 6-11, Aug. 2011.

[30] M. J. Deen, "Information and communications technologies for elderly ubiquitous healthcare in a smart home," *Pers. Ubiquitous Comput.*, vol. 19, no. 3, pp. 573-99, June 2015.

[31] J. Kim, H. Choi, H. Wang, N. Agoulmine, M. J. Deerv, and J. W. Hong, "POSTECH's U-Health Smart Home for elderly monitoring and support," *2010 IEEE International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM)*, pp. 14-17, June 2020.

[32] S. Majumder, T. Mondal, and M. J. Deen, "Wearable sensors for remote health monitoring," *Sensors (Basel)*, vol. 17, no. 1, p. 130, Jan. 2017.

[33] S. Intille, K. Larson, E. M. Tapia, J. S. Beaudin, P. Kaushik, J. Nawyn, and R. Rockinson, "Using a live-in laboratory for ubiquitous computing research," In *Lecture Notes in Computer Science*, K. P. Fishkin, B. Schiele, P. Nixon, A. Quigley. Berlin, Germany: Springer-Verlag, 2006, vol. 3968, pp. 349-365.

[34] M. J. Rantz, M. Skubic, and S. J. Miller, "Using sensor technology to augment traditional health-care," in *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, 2009, pp. 6159–6162.

[35] J. A. Williams, and D. J. Cook, “Forecasting behavior in smart homes based on sleep and wake patterns,” *Technol. Health Care*, vol. 25, no. 2, pp. 89-110, 2017.

[36] P. G. Jacobs and J. A. Kaye, “Ubiquitous real-world sensing and audiology-based health informatics,” *J. Am. Acad. Audiol.*, vol. 26, no. 9, pp. 777–783, Oct. 2015.

[37] C. VandeWeerd, A. Yalcin, G. Aden-Buie, Y. Wang, M. Roberts, N. Mahser, C. Fnu, and D. Fabiano, “HomeSense: Design of an ambient home health and wellness monitoring platform for older adults,” *Health and Technol.*, accepted for publication.

[38] Safer, smarter homes start with Z-wave. [Online]. Available: http://www.z-wave.com/

[39] What is MQTT. [Online]. Available: http://mqtt.org/

[40] Y. Wang, A. Yalcin, and C. VandeWeerd, “Health and wellness monitoring using ambient sensor networks,” *J. Amb. Intel. Smart En.*, vol. 12, no. 2, pp. 139-151, March 2020. doi: 10.3233/AIS-200553

[41] Forum on Aging-Releases Statistics, “Older Americans 2016: Key indicators of well-being,” *Federal Interagency Forum on Aging-Related Statistics*. Washington, DC, USA, Government Printing Office, Aug. 2016. [Online]. Available: https://agingstats.gov/docs/LatestReport/Older-Americans-2016-Key-Indicators-of-WellBeing.pdf

[42] Eldercare Workforce Alliance, “Caring for an aging America: Meeting the health care needs of older adults,” Washington, DC, USA, Nov. 2013. [Online]. Available: https://eldercareworkforce.org/files/QA_Issue_Brief_-_FINAL.pdf

[43] S. C. Reinhard, L. F. Feinberg, R. Choula, and A. Houser, “Valuing the invaluable: 2015 update. Undeniable progress, but big gaps remain,” *AARP Public Policy Institute*, Washington, DC, USA, July 2015. [Online]. Available: https://www.aarp.org/content/dam/aarp/ppi/2015/valuing-the-invaluable-2015-update-new.pdf

[44] S. C. Reinhard, B. Given, N. H. Petlick, and A. Bemis, “Supporting family caregivers in providing care,” in *Patient safety and quality: An evidence-based handbook for nurses*, Agency for Healthcare Research and Quality, R. G. Hughes, Ed., Rockville, MD, USA, Apr. 2008. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK2665/

[45] D. J. Cook, J. C. Augusto, and V. R. Jakkula. “Ambient intelligence: technologies, applications, and opportunities,” *Pervasive and Mob. Comput.*, vol. 5, no. 4, pp. 227-298, Aug. 2009.

[46] F. Sadri, “Ambient intelligence: a survey,” *ACM Comput. Surv.*, vol. 43, no. 4, no. 36, pp. 36:1-36:66, Oct. 2011.

[47] G. Acampora, D. J. Cook, P. Rashidi, and A. V. Vasilakos, "A survey on ambient intelligence in healthcare," *Proc. IEEE*, vol. 101, no. 12, pp. 2470-2494, Dec. 2013.

[48] P. Rashidi and A. Mihailidis, "A survey on ambient-assisted living tools for older adults," IEEE J. Biomed. *Health Inform.*, vol. 17, no. 3, pp. 579-590, May 2013.

[49] National Council on Aging, "The United States of Aging Survey Executive Summary," Arlington, VA, July 2015.

[50] D. J. Cook, M. Youngblood, E. O. Heierman, K. Gopalratnam, S. Rao, A. Litvin, and F. Khawaja, "MavHome: An agent-based smart home," in *Proc. 1st IEEE Int. Conf. Perv. Comput. Commun.*, Fort Worth, TX, May 2003, pp. 521-524.

[51] C. D. Kidd, R. Orr, G. D. Abowd, C. G. Atkeson, I. A. Essa, B. MacIntyre, E. Mynatt, T. E. Starner, and W. Newstetter, "The Aware Home: A living laboratory for ubiquitous computing research," in *Proc. 2nd Int. Workshop Cooperative Buildings*, Pittsburgh, PA, USA, Oct. 1999, pp. 191-198.

[52] D. J. Cook, A. S. Crandall, B. L. Thomas, and N. C. Krishnan, "CASAS: a smart home in a box," *Computer*, vol. 46, no. 7, pp. 62-69, July 2013.

[53] M. J. Rantz, K. D. Marek, M. A. Aud, R. A. Johnson, D. Otto, and R. Porter, "TigerPlace: A new future for older adults," *J. Nursing Care Quality*, vol. 20, no. 1, pp. 1-4, 2005.

[54] J. A. Kaye, S. A. Maxwell, N. Mattek, T. L. Hayes, H. Dodge, M. Pavel, H. B. Jimison, K. Wild, L. Boise, and T. A. Zizelberger, "Intelligent systems for assessing aging changes: Home-based, unobtrusive, and continuous assessment of aging," *J. Gerontol. B Psychol. Sci. Soc. Sci.*, vol. 66B, no. 1, pp. i180-i190, July 2011.

[55] J. Doyle, A. Kealy, J. Loane, L. Walsh, B. O'Mullane, C. Flynn, R. Bond, A. Macfarlane, B. Bortz, and R. B. Knapp, "An integrated home-based self-management system to support the wellbeing of older adults," *J. Ambient Intell. Smart Environ.*, vol. 6, no. 4, pp. 359-383, 2014.

[56] S. Ohta, H. Nakamoto, Y. Shinagawa, and T. Tanikawa, "A health monitoring system for elderly people living alone," *J. Telemed. Telecare*, vol. 8, no. 3, pp. 151-156, June 2002.

[57] G. Virone, M. Alwan, S. Dalal, S. W. Kell, B. Turner, J. A. Stankovic, and R. Felder, "Behavioral patterns of older adults in assisted living," *IEEE Trans. Inf. Technol. Biomed.*, vol. 12, no. 3, pp.387-398, May 2008..

[58] S. Wang, M. Skubic, and Y. Zhu, "Activity density map visualization and dissimilarity comparison for eldercare monitoring," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 4, pp. 607-614, July 2012.

[59] A. O'Brien, K. McDaid, J. Loane, J. Doyle, and B. O'Mullane, "Visualization of movement of older adults within their Homes based on PIR sensor data," in *Proc. 6th Int. Conf. Pervasive Comput. Technol. Healthc.*, San Diego, CA, USA, May 2012, pp. 252-259.

[60] F. Palumbo, D. La Rosa, and E. Ferro, "Stigmergy-based long-term monitoring of indoor users mobility in ambient assisted living environments: the DOREMI project approach," in: *Proc. 2nd Italian Workshop on Artificial Intelligence for Ambient Assisted Living (AI*AAL 2016)*, vol. 1803, pp. 18-32, 2016.

[61] I. Susnea, L. Dumitriu, M. Talmaciu, E. Pecheanu, and D. Munteanu, "Unobtrusive monitoring the daily activity routine of elderly people living alone, with low-cost binary sensors," *Sensors*, vol. 19, no. 10, May 2019.

[62] I. Susnea, "Engineering human stigmergy," *Int. J. Comput. Commun.*, vol. 10, no. 3, pp. 420–427, Apr. 2015.

[63] A. Jain, M. Popescu, J. Keller, M. Ranta, and B. Markway, "Linguistic summarization of in-home sensor data," *J. Biomed. Inform.*, vol. 96, pp. 1-14, Aug. 2019.

[64] M. P. Lawton and E. M. Brody, "Assessment of older people: Self-Maintaining and instrumental activities of daily living," *The Gerontologist*, vol. 9, no. 3, pp. 179-186, Oct. 1969.

[65] D. Foti and J. S. Koketsu, "Activities of daily living," in *Pedretti's Occupational Therapy: Practical Skills for Physical Dysfunction*, Elsevier Health Sciences: Amsterdam, Netherlands, 2013, vol. 7, pp. 157–232.

[66] J. Petersen, D. Austin, J. A. Kaye, M. Pavel, and T. L. Hayes, "Unobtrusive in-home detection of time spent out-of-home with applications to loneliness and physical activity," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 5, pp 1590-1596, Sep. 2014.

[67] D. J. Cook and M. Schmitter-Edgecombe, "Assessing the quality of activities in a smart environment," *Method Inform. Med.*, vol. 48, no. 5, pp. 480-485, May 2009.

[68] P. N. Dawadi, D. J. Cook, and M. Schmitter-Edgecombe, "Automated cognitive health assessment from smart home-based behavior data," *IEEE J. Biomed. Health Inform.*, vol. 20, no. 4, pp. 1188-1194, July 2016.

[69] A. A. Aramendi, A. Weakley, A. A. Goenaga, M. Schmitter-Edgecombe, and D. J. Cook, "Automatic assessment of functional health decline in older adults based on smart home data," *J. Biomed. Inform.*, vol. 81, pp. 119-130, May 2018.

[70] A. Alberdi, A. Weakley, M. Schmitter-Edgecombe, and D. J. Cook, "Smart home-based prediction of multidomain symptoms related to Alzheimer's disease," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 6, pp. 2168-2194, Nov. 2018.

[71] N. C. Krishnan and D. J. Cook, “Activity recognition on streaming sensor data,” *Pervasive Mob. Comput.*, vol. 10, pp. 138-154, Jan. 2014.

[72] A. Ghods, K. Caffrey, B. Lin, K. Fraga, R. Fritz, M. Schmitter-Edgecombe, C. Hundhausen, and D. J. Cook, “Iterative design of visual analytics for a clinician-in-the-loop smart home,” *IEEE J. Biomed. Health Inform.*, vol. 23, no. 4, pp. 1742-1748, July 2019.

[73] T. L. Hayes, T. Riley, M. Pavel, and J. A. Kaye, “Estimation of rest-activity patterns using motion sensors,” in *Proc. Conf. Proc. IEEE Eng. Med. Biol. Soc.*, 2010, pp. 2147-2150.

[74] A. Kealy, K. McDaid, J. Loane, L. Walsh, and J. Doyle, “Derivation of night time behaviour metrics using ambient sensors,” in *Proc. 7th Int. Conf. Pervasive Comput. Technol. Healthc.*, Venice, Italy, 2013, pp. 33-40.

[75] T. Hayes, M. Pavel, and J. Kaye, “An approach for deriving continuous health assessment indicators from in-home sensor data,” in *Proc. Technol. Aging: Select. Papers Int. Conf. Technol. Aging*, vol. 21, pp. 130-137, Dec. 2008.

[76] N. Goonawardene, X. Toh, and H. Tan, “Sensor-driven detection of social isolation in community-dwelling elderly,” in *Human Aspects of IT for the Ages Population (ITAP 2017). Applications, Services and Context,* May 2017, pp. 378-392.

[77] Y. EI-Khadiri, G. Corona, C. Rose, and F. Charpillet, “Sleep activity recognition using binary motion sensors,” in *2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI 2018)* , Volos, Greece, pp. 265-269.

[78] D. Arifoglu and A. Bouchachia, “Activity recognition and abnormal behaviour detection with recurrent neural networks,” *Procedia Comput. Sci.*, vol. 110, pp. 86-93, Jul. 2017.

[79] H. H. Dodge, N. C. Mattek, , T. L. Hayes, and J. A. Kaye, “In-home walking speeds and variability trajectories associated with mild cognitive impairment,” *Neurology*, vol. 78, no. 24, pp. 1945-1952, Jun. 2012.

[80] T. Banerjee, M. Yefimova, J. M. Keller, M. Skubic, D. L. Woods, and M. Rantz, “Exploratory analysis of older adults’ sedentary behavior in the primary living area using kinect depth data,” *J. Amb. Intel. Smart En.*, vol. 9, no. 2, pp. 163-179, Feb. 2017.

[81] N. Vaney, A. Dixit, T. Ghosh, R. Gupta, M. S. Bhatia, “Habituation of event related potentials: a tool for assessment of cognition in headache patients,” *J. Delhi Psychiatry*, vol. 11, no. 1, Apr. 2008.

[82] D. Dodou, J. C. F. de Winter, “Social desirability is the same in offline, online and paper surveys: A meta-analysis,” *Comput. Human Behav.*, vol. 36, pp. 487-495, July 2014.

[83] H. Barbosa, M. Barthelemy, G. Ghoshal, C. R. James, M. Lenormand, T. Louail, R. Menezesh, Jose J. Ramasco, F. Simini, and M. Tomasini, “Human mobility: models and applications,” *Physics Reports*, vol. 734, no. 6, pp. 1-74, Mar. 2018.

[84] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, “Limits of predictability in human mobility,” *Science*, vol. 327, no. 5968, pp. 1018-1021, Feb. 2010.

[85] X. Lu, E. Wetter, N. Bharti, A. J. Tatem, and L. Bengtsson, “Approaching the limit of predictability in human mobility,” *Sci. Rep.*, vol. 3, no. 10, pp. 1-9, Oct. 2013.

[86] Y. Zhang, Q. Li, Y. Chen, X. Xie, and W. Ma, “Understanding mobility based on GPS data,” In *Proc. 10th Int. Conf. Ubiquitous Computing. (UbiComp’08)*, Seoul, Korea, Sept. 2008.

[87] E. L. Ikanovic and A. Mollgaard, “An alternative approach to the limits of predictability in human mobility,” *EPJ Data Sci.*, vol. 6, no. 12, June 2017.

[88] A. Cuttone, S. Lehmann, and M. C. Gonzalez, “Understanding predictability and exploration in human mobility,” *EPJ Data Sci.*, vol. 7, no. 2, Jan. 2018.

[89] P. Y. Cao, G. Li, A. C. Champion, D. Xuan, S. Romig, and W. Zhao, "On human mobility predictability via WLAN logs," in *Proc. IEEE Comput. Commun. (IEEE INFOCOM 2017)*, May 2017, pp. 1-9.

[90] G. Goulet-Langlois, H. N. Koutsopoulos, Z. Zhao, and J. Zhao, “Measuring regularity of individual travel patterns,” *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 5, pp. 1583-1592, May 2018.

[91] G. Qi, X. Li, S. Li, G. Pan, Z. Wang, and D. Zhang, "Measuring social functions of city regions from large-scale taxi behaviors," in *2011 IEEE Intel. Conf. Pervasive Comput. Commun. Workshops (PERCOM Workshops)*, pp. 384-388.

[92] Y. Ge, H. Xiong, A. Tuzhilin, K. Xiao, M. Gruteser, and M. Pazzani, “An energy-efficient mobile recommender system,” in *Proc. 16th ACM SIGKDD Intel. Conf.*, 2010, pp. 899-908.

[93] V. Colizza, A. Barrat, M. Barthélemy, A. Vespignani, “Predictability and epidemic pathways in global outbreaks of infectious diseases: the SARS case study,” *BMC Med.*, p. 5-34, Nov. 2007.

[94] V. Belik, T. Geisel, and D. Brockmann, “Natural human mobility patterns and spatial spread of infectious diseases,” *Phys. Rev. X*, vol. 1, no. 1, pp. 1-11, Aug. 2011.

[95] L. Bengtsson, X. Lu, A. Thorson, R. Garfield, and J. von Schreeb, “Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: A post-earthquake geospatial study in Haiti,” *PLos Medicine*, vol. 8, no. 8, pp. 1-9, Aug. 2011.

[96] M. Stute, M. Maass, T. Schons, and M. Hollick, “Reverse engineering human mobility in large-scale natural disasters,” *In Proc. MSWiM ’17*, pp. 219-226.

[97] S. K. Das, D. J. Cook, A. Battacharya, E. O. Heierman, and Tze-Yun Lin, "The role of prediction algorithms in the MavHome smart home architecture," in *IEEE Wireless Communications Magazine*, vol. 9, no. 6, pp. 77-84, Dec. 2002.

[98] M. C. Mozer, R. H. Dodier, M. Anderson, L. Vidmar, R. F. Cruickshank, and D. Miller, “The neural network house: An environment that adapts to its inhabitants,” in *Proc. AAAI Spring Symp. Intell. Environ.*, 1998, pp. 110-114.

[99] K. Bao, F. Allerding, and H. Schmeck, “User behavior prediction for energy management in smart homes,” in *Proc. IEEE FSKD’11*, pp. 1335-1339, July 2011.

[100] M. Chan, C. Hariton, P. Ringeard, and E. Campo, “Smart house automation system for the elderly and the disabled,” in *Proc. IEEE Int. Conf. Systems, Man and Cybernetics*, Vancouver, BC, Canada, 1995, pp. 1586-1589.

[101] S. K. Das and D. J. Cook, “Health monitoring in an agent-based smart home by activity prediction,” in *Proc. IEEE Int. Conf. Smart Homes and Health Telematics (ICOST)*, Singapore, vol. 14, pp. 3-14, Sept. 2004.

[102] M. Chan, S. Bonhomme, D. Esteve, and E. Campo, “Individual movement trajectories in smart homes,” *ICBME*, 2008, pp. 1014-1018.

[103] J. Hao, B. Bouchard, A. Bouzouane and S. Gaboury, "Real-time activity prediction and recognition in smart homes by formal concept analysis," *2016 12th International Conference on Intelligent Environments (IE)*, London, 2016, pp. 103-110.

[104] M. R. Alam, M. B. I. Reaz, and M. A. M. Ali, “A review of smart homes - Past, present, and future,” *IEEE Trans. Syst., Man, Cybern. C*, vol. 42, no. 6, pp. 1190–1203, Nov. 2012.

[105] P. Remagnino and G. L. Foresti, “Ambieint intelligence: A new Mmltidisciplinary paradigm,” *IEEE Trans. Syst., Man, Cybern. A*, vol. 35, no. 1, pp. 1–6, Jan. 2005.

[106] D. Austin, R. M. Cross, T. Hayes, and J. Kaye, “Regularity and predictability of human mobility in personal space,” *PLoS ONE*, vol. 9, no. 2, p. e90256, 2014.

[107] Y.S. Niu, N. Hao, and H. Zhang, “Multiple change-point detection: a selective overview,” *Sci.*, vol. 31, no. 4, pp. 611-623, July 2016.

[108] C. Truong, L. Oudre, and N. Vayatis, “Selective review of offline change point detection methods,” *Signal Process.*, 167:107299, Feb. 2020.

[109] G. Schwarz, “Estimating the dimension of a model,” *Ann. Statist.*, vol. 6, no. 2, pp. 461-464, 1978.

[110] Y. C. Yao, "Estimating the number of change-points via Schwarz's criterion," *Stat. Probabil. Lett.*, vol. 6, pp. 181-189, Feb. 1988.

[111] M. Lavielle, "Detection of multiple changes in a sequence of dependent variables," *Stoch. Proc. Appl.*, vol. 83, pp. 79-102, Sept. 1999.

[112] M. Lavielle and E. Moulines, "Least-squares estimation of an unknown number of shifts in a time series," *J. Time Ser. Anal.*, vol. 21, no. 1, pp. 33-59, Jan. 2000.

[113] F. Barraquand and S. Benhamou, "Animal movements in heterogeneous landscapes identifying profitable places and homogeneous movement bouts," *Ecology*, vol. 89, no. 12, pp. 3336-3348, Dec. 2008.

[114] M. Lavielle, "Using penalized contrasts for the change-point problem," *Signal Process.*, vol. 85, pp. 1501–1510, 2005.

[115] F. Picard, S. Robin, M. Lavielle, C. Vaisse, and J. J. Daudin, "A statistical approach for array CGH data analysis," *BMC Bioinformatics*, vol. 6, no. 27, pp. 1-14, Aug. 2005.

[116] J. Gazeaux, S. Williams, M. King, M. Bos, R. Dach, M. Deo, et al., "Detecting offsets in GPS time series: First results from the detection of offsets in GPS experiment," *J. Geophys. Res. Solid Earch*, vol. 118, pp. 2397–2407, Mar. 2013.

[117] T. M. Cover and J. A. Thomas, "Entropy rates of a stochastic process," in *Elements of Information Theory*, 2nd ed. New York, Wiley, 1991, Chapter 4, pp. 63-65.

[118] H. Cai, S. R. Kulkarni, and S. Verdu, "Universal entropy estimation via block sorting," *IEEE Trans. Inf. Theory*, vol. 50, no. 7, pp. 1551-1561, Jul. 2004.

[119] G. Strang*, Calculus,* New York: Wellesley-Cambridge Press, 1991.

[120] M. Sur, A. K. Skidmore, K. M. Exo, T. Wang, B. J. Ens, and A. G. Toxopeus, "Change detection in animal movement using discrete wavelet analysis," *Ecol. Inform.*, vol. 20, pp. 47-57, Mar. 2014.

[121] C. Calenge, "Analysis of animal movements in R: the adehabitatLT package," in Office national de la chasse et de la faune sauvage, Saint Benoist, France, Apr. 2011.

[122] B. L. Welch, "The generalization of "Student's" problem when several different population variances are involved," *Biometrika*, vol. 34, no. 1/2, pp. 28–35, Jan. 1947.

[123] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. Classification and regression trees. Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software, 1984.

[124] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover's distance as a metric for image retrieval," *Int. J. Comput. Vision*, 40(2)99-121, Nov. 2000.

[125] I. J. Myung, “Tutorial on maximum likelihood estimation,” *J. Math. Psychol.*, vol. 47, no. 1, pp. 90–100, Feb. 2003.

[126] M. Basseville and I. Nikiforov, Detection of abrupt changes: Theory and application. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[127] S. Aminikhanghahi and D. J. Cook, “A Survey of methods for time series change point detection.” *Knowl Inf Syst.*, vol 51, no 2, pp. 339-367, May 2017.

[128] J. Chen and A. K. Gupta, “On change point detection and estimation”, *Comm. Statistics – Simulation and Computation*, vol. 30, no. 3, pp. 665-697, Feb. 2007.

Appendix A: Copyright Permission from Springer

The permission below is for the reuse of the Introduction of the paper "HomeSense: Design of an Ambient Home Health and Wellness Monitoring Platform for Older Adults" accepted by *Health and Technology* in Chapter 1.

Copyright Transfer Statement

The copyright to this article is transferred to Systems Engineering Society of China and Springer (respective to owner if other than Systems Engineering Society of China and Springer and for U.S. government employees: to the extent transferable) effective if and when the article is accepted for publication. The author warrants that his/her contribution is original and that he/she has full power to make this grant. The author signs for and accepts responsibility for releasing this material on behalf of any and all co-authors. The copyright transfer covers the exclusive right and license to reproduce, publish, distribute and archive the article in all forms and media of expression now known or developed in the future, including reprints, translations, photographic reproductions, microform, electronic form (offline, online) or any other reproductions of similar nature.

An author may self-archive an author-created version of his/her article on his/her own website and or in his/her institutional repository. He/she may also deposit this version on his/her funder's or funder's designated repository at the funder's request or as a result of a legal obligation, provided it is not made publicly available until 12 months after official publication. He/ she may not use the publisher's PDF version, which is posted on www.springerlink.com, for the purpose of self-archiving or deposit. Furthermore, the author may only post his/her version provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at www.springerlink.com".

Prior versions of the article published on non-commercial pre-print servers like arXiv.org can remain on these servers and/or can be updated with the author's accepted version. The final published version (in pdf or html/xml format) cannot be used for this purpose. Acknowledgement needs to be given to the final publication and a link should be inserted to the published article on Springer's website, accompanied by the text "The final publication is available at springerlink.com".

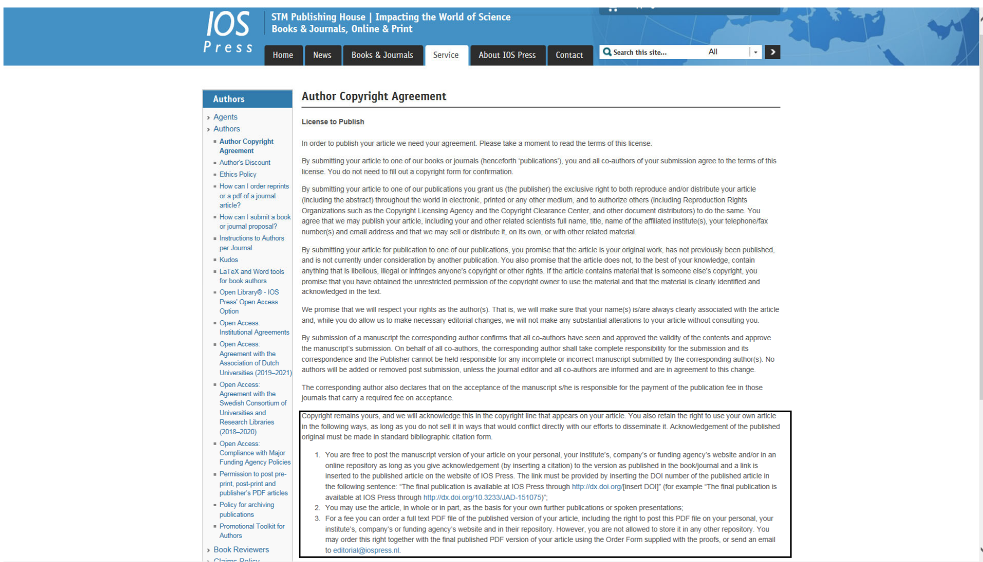
The author retains the right to use his/her article for his/her further scientific career by including the final published journal article in other publications such as dissertations and postdoctoral qualifications provided acknowledgement is given to the original source of publication.

The author is requested to use the appropriate DOI for the article. Articles disseminated via www.springerlink.com are indexed, abstracted and referenced by many abstracting and information services, bibliographic networks, subscription agencies, library networks, and consortia.

After submission of the agreement signed by the corresponding author, changes of authorship or in the order of the authors listed will not be accepted by Systems Engineering Society of China and Springer.

Appendix B: Copyright Permission from IOS Press

This permission below is for the reuse of the paper “Health and wellness monitoring based on ambient networks” published by *Journal of Ambient Intelligence and Smart Environments* in Chapter 2.



IOS Press

STM Publishing House | Impacting the World of Science

Books & Journals, Online & Print

Home | News | Books & Journals | Service | About IOS Press | Contact

Search this site... All

Authors

- Agents
- Authors
  - **Author Copyright Agreement**
  - Author's Discount
  - Ethics Policy
  - How can I order reprints or a pdf of a journal article?
  - How can I submit a book or journal proposal?
  - Instructions to Authors per Journal
  - Kudos
  - LaTeX and Word tools for book authors
  - Open Library® - IOS Press' Open Access Option
  - Open Access: Institutional Agreements
  - Open Access: Agreement with the Association of Dutch Universities (2019–2021)
  - Open Access: Agreement with the Swedish Consortium of Universities and Research Libraries (2018–2020)
  - Open Access: Compliance with Major Funding Agency Policies
  - Permission to post pre-print, post-print and publisher's PDF articles
  - Policy for archiving publications
  - Promotional Toolkit for Authors
- Book Reviewers
- Claims Policy

Author Copyright Agreement

License to Publish

In order to publish your article we need your agreement. Please take a moment to read the terms of this license.

By submitting your article to one of our books or journals (henceforth ‘publications’), you and all co-authors of your submission agree to the terms of this license. You do not need to fill out a copyright form for confirmation.

By submitting your article to one of our publications you grant us (the publisher) the exclusive right to both reproduce and/or distribute your article (including the abstract) throughout the world in electronic, printed or any other medium, and to authorize others (including Reproduction Rights Organizations such as the Copyright Licensing Agency and the Copyright Clearance Center, and other document distributors) to do the same. You agree that we may publish your article, including your and other related scientists full name, title, name of the affiliated institute(s), your telephone/fax number(s) and email address and that we may sell or distribute it, on its own, or with other related material.

By submitting your article for publication to one of our publications, you promise that the article is your original work, has not previously been published, and is not currently under consideration by another publication. You also promise that the article does not, to the best of your knowledge, contain anything that is libellous, illegal or infringes anyone’s copyright or other rights. If the article contains material that is someone else’s copyright, you promise that you have obtained the unrestricted permission of the copyright owner to use the material and that the material is clearly identified and acknowledged in the text.

We promise that we will respect your rights as the author(s). That is, we will make sure that your name(s) is/are always clearly associated with the article and, while you do allow us to make necessary editorial changes, we will not make any substantial alterations to your article without consulting you.

By submission of a manuscript the corresponding author confirms that all co-authors have seen and approved the validity of the contents and approve the manuscript’s submission. On behalf of all co-authors, the corresponding author shall take complete responsibility for the submission and its correspondence and the Publisher cannot be held responsible for any incomplete or incorrect manuscript submitted by the corresponding author(s). No authors will be added or removed post submission, unless the journal editor and all co-authors are informed and are in agreement to this change.

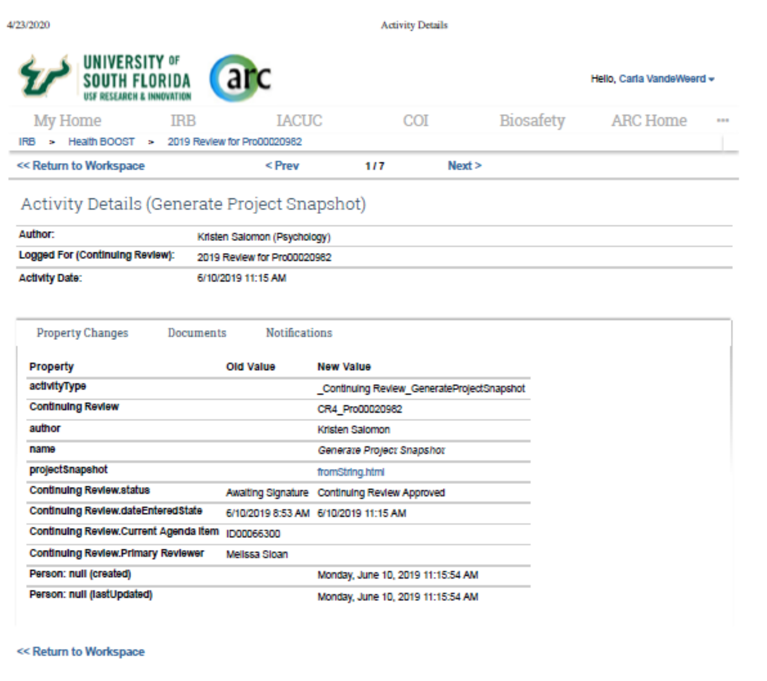
The corresponding author also declares that on the acceptance of the manuscript s/he is responsible for the payment of the publication fee in those journals that carry a required fee on acceptance.

Copyright remains yours, and we will acknowledge this in the copyright line that appears on your article. You also retain the right to use your own article in the following ways, as long as you do not sell it in ways that would conflict directly with our efforts to disseminate it. Acknowledgement of the published original must be made in standard bibliographic citation form.

1. You are free to post the manuscript version of your article on your personal, your institute’s, company’s or funding agency’s website and/or in an online repository as long as you give acknowledgement (by inserting a citation) to the version as published in the book/journal and a link is inserted to the published article on the website of IOS Press. The link must be provided by inserting the DOI number of the published article in the following sentence: “The final publication is available at IOS Press through http://dx.doi.org/[insert DOI]” (for example “The final publication is available at IOS Press through http://dx.doi.org/10.3233/JAD-151075)”;
2. You may use the article, in whole or in part, as the basis for your own further publications or spoken presentations;
3. For a fee you can order a full text PDF file of the published version of your article, including the right to post this PDF file on your personal, your institute’s, company’s or funding agency’s website and in their repository. However, you are not allowed to store it in any other repository. You may order this right together with the final published PDF version of your article using the Order Form supplied with the proofs, or send an email to editorial@iospress.nl.

Appendix C: IRB Review Approval

The picture below indicates that PRO 00020982 is approved by IRB.



4/23/2020 Activity Details

UNIVERSITY OF SOUTH FLORIDA USF RESEARCH & INNOVATION arc

Hello, Carla VandeWeerd

My Home IRB IACUC COI Biosafety ARC Home

IRB > Health BOOST > 2019 Review for Pro00020982

<< Return to Workspace < Prev 1 / 7 Next >

Activity Details (Generate Project Snapshot)

Author:	Kristen Salomon (Psychology)
Logged For (Continuing Review):	2019 Review for Pro00020982
Activity Date:	6/10/2019 11:15 AM

Property Changes Documents Notifications

Property	Old Value	New Value
activityType		_Continuing Review_GenerateProjectSnapshot
Continuing Review		CR4_Pro00020982
author		Kristen Salomon
name		Generate Project Snapshot
projectSnapshot		fromString.html
Continuing Review.status	Awaiting Signature	Continuing Review Approved
Continuing Review.dateEnteredState	6/10/2019 8:53 AM	6/10/2019 11:15 AM
Continuing Review.Current Agenda Item	ID00066300	
Continuing Review.Primary Reviewer	Melissa Sloan	
Person: null (created)		Monday, June 10, 2019 11:15:54 AM
Person: null (lastUpdated)		Monday, June 10, 2019 11:15:54 AM

<< Return to Workspace

Figure A. IRB review approval.