Identifying change in human behaviour utilising static sensors

Jess Gillam, B.Sc. (Hons.), M.Res

Submitted for the degree of Doctor of
Philosophy at Lancaster University.

July 2022

STOR-i
excellence with impact

Abstract

By 2039, it is projected that one in four people will be aged 65 or over. It is important we look for new ways to help and care for an ageing population. This thesis introduces new methods for identifying changes in behaviour within households, using observations from passive sensors placed within the homes of older people. First, we propose a novel method for detecting subtle changes in sequences whilst taking into account the natural day-to-day variability and differing numbers of 'trigger' events per day. Next we introduce a model to predict the probability a sensor will trigger throughout the day for a household, whilst considering the prior data and other sensors within the home. Finally, we present a framework to identify changes in probabilistic cluster membership of households across time. We assess the performance for each of these methods on simulated data and data provided through our partnership with Howz.

Acknowledgements

First of all, I am so eternally grateful to my supervisor Rebecca Killick. This absolutely wouldn't have been possible without your amazing support and guidance. I have learnt so much from you and I am going to be so sad not seeing you every week.

Next, I would like to thank everyone involved with STOR-i Centre for Doctoral Training for all the opportunities they have given me and the friendly atmosphere they provided over the years. In particular Jonathan Tawn, your kindness and care for every student does not go unnoticed!

Thank you to Jonathan Burr, Louise Rogerson and all of Howz for the support you have given me, the data provided and the insight into your wonderful company. In particular, thank you to my industrial supervisor Ben Norwood for the enthusiastic discussions about my work and the company.

To my family, thank you for your endless love and support. I appreciate you all so much. To my partner, Mike, thank you for always being by my side and making me smile, even on my worst days. Thank you to Charley, Mirjam and Lisa for always checking in on me and supporting me through all the ups and downs. Finally, thank you to the Tigersharks for all the laughter and friendship over the last few years. I couldn't have gotten here without all of you!

Declaration

I declare that the work in this thesis has been done by myself and has not been submitted elsewhere for the award of any other degree.

Jess Gillam

List of Papers

This thesis includes the following papers:

- Gillam, Killick, Taylor, Heal, and Norwood (2022+b), *Identifying Irregular Activity Sequences: an Application to Passive Household Monitoring*, Submitted.
- Gillam, Killick, Heal, and Norwood (2022), *Modelling and Forecasting of at Home Activity in Older Adults using Passive Sensor Technology*, Statistics in Medicine, *In Press.*
- Gillam, Killick, and Norwood (2022+a), *Framework to Identify Subtle Changes in Probabilistic Cluster Membership: an application to passive household monitoring*, Submitted.

These papers are given as Chapters 2, 3 and 4 respectively. Each paper contains its own motivation and review of the appropriate literature. The notation is introduced in each chapter, allowing them to be read as a whole or individually. The appendices and bibliography can be found at the end of the thesis.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

It is projected that one in four people in the UK will be aged 65 or over by 2039 (ONS, 2021), due to increased life expectancy and other factors (Age UK, 2019). As of April 2019, there are just under 12 million people in the UK aged 65 or over, with close to a quarter of those living alone. They constitute 42% of elective admissions and 43% of emergency admissions to hospital (Age UK, 2019). Due to this, new ways to help the ageing population and reduce the strain on health care services are becoming more necessary.

Increasingly, people are turning to technology to help improve health and care of the elderly. One avenue is placing technology within the home and using the data it provides. Devices used include; medical devices measuring vitals - blood pressure and heart rate monitoring, wearable technology - accelerometers and pedometers, and smart home sensors such a motion sensors, appliance sensors and door sensors which trigger an observation when activated by the user. Mobile applications have also been developed to compile the data each of the devices collect and incorporate information already taken from other health applications; e.g. pedometer.

Depending on the type of care, the output from devices are used for different purposes. For example; individual households, where the user lives independently within their own home. The information provided by the devices could go to the user itself, carer or family members to provide peace of mind on the well-being of the household. Furthermore, alerts can be sent to the clients, when changes in the device data are found. For carers and health care professionals, this results in the identification of issues in well-being as quickly as possible. The benefits here are quick interventions to prevent hospitalisation and a reduction in pressure on GP services. This system of devices can also be used to monitor symptoms of long-term conditions and illnesses, e.g. dementia, diabetes, depression and anxiety. Allowing carers to adapt the care plan quickly to the changing needs of the user and identifying when intervention is necessary.

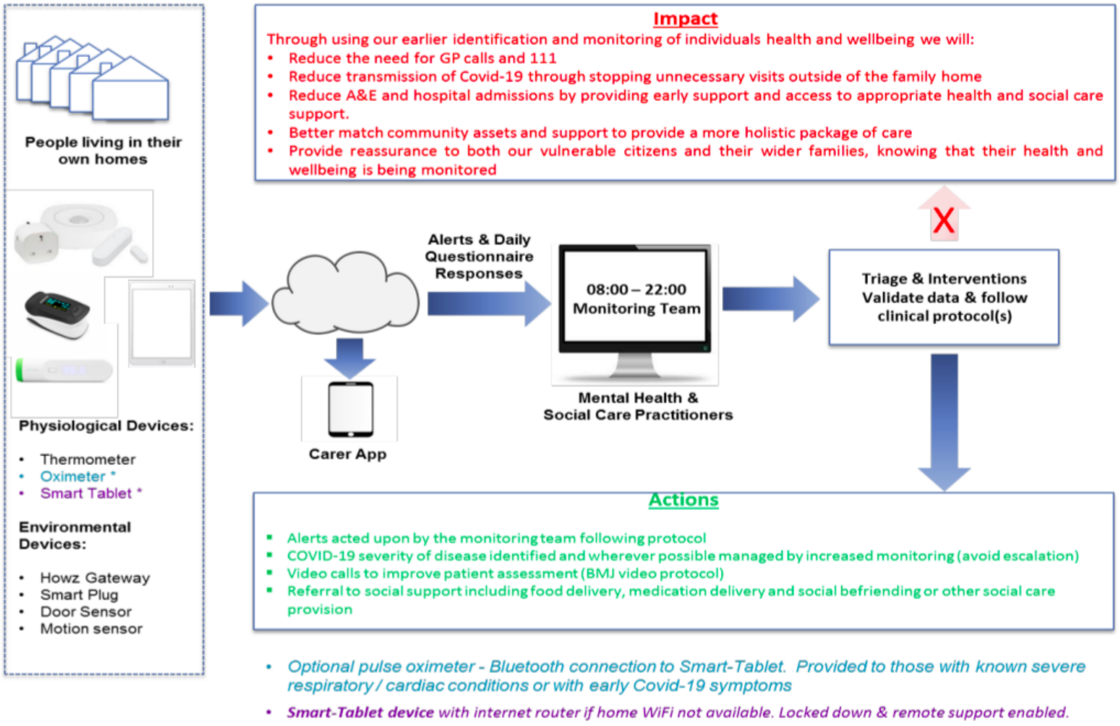
The data collected from these devices can also be provided to support the National Health Service (NHS) for the UK. The medical devices are able to provide remote measuring of vitals so that monitoring a patients' condition from their own home is more accessible. This allows for earlier release from the hospital, reducing hospital stays and pressure on bed space. Further, reassurance given by the system of devices could reduce unnecessary emergency ambulance callouts.

Using a combination of these devices allows the clients and healthcare professionals to understand the daily routines and behaviour of the household. Furthermore, in order to support the household to live independently for as long as possible, the devices can monitor and help to identify changes in the behaviour of households so that prompt intervention can occur.

Howz is a product that can provide such a system of sensors. Howz introduce a number of non-intrusive passive sensors within the homes of elderly and vulnerable people and a mobile app for the user to see insights from their data. In a recent research study with the NHS, the Howz product has been implemented in homes to show the potential benefits of such a service in the care of Dementia patients (KSS et al., 2021). Dementia is a term for a group of symptoms with various diseases being the cause e.g., Alzheimer's. There exists various literature studying dementia and the prevalence in the population (Prince et al., 2014). The most common cause of dementia is Alzheimer's disease with some of the following symptoms; early stage - difficulty remembering recent conversations, names or events and later stage symptoms - impaired communication, disorientation, behavioural changes, difficulty swallowing and talking (Alzheimer's Association, 2020). Monitoring users with dementia through a system of sensors has the potential to identify when a user progresses into the later stages, as well as identifying new symptoms and changes quickly so that carers and health care professionals can support the user however they need it.

Figure 1.0.1 shows a visual representation of how the Howz sensors have been used in this study. Howz provide the following pathway for this system; placing the sensors within the homes of the patients, monitoring the data and if something of concern is detected the patient is provided with triage and/or intervention. Alerts are sent when the results of data analysis show the patient is acting outside of their 'normal' level, established during the first seven days of using the sensors.

Using the Howz product in the study has shown a reduction of emergency admis-



Through using our earlier identification and monitoring of individuals health and wellbeing we will:

- Reduce the need for GP calls and 111
- Reduce transmission of Covid-19 through stopping unnecessary visits outside of the family home
- Reduce A&E and hospital admissions by providing early support and access to appropriate health and social care support.
- Better match community assets and support to provide a more holistic package of care
- Provide reassurance to both our vulnerable citizens and their wider families, knowing that their health and wellbeing is being monitored

- Alerts acted upon by the monitoring team following protocol
- COVID-19 severity of disease identified and wherever possible managed by increased monitoring (avoid escalation)
- Video calls to improve patient assessment (BMJ video protocol)
- Referral to social support including food delivery, medication delivery and social befriending or other social care provision

- *Optional pulse oximeter - Bluetooth connection to Smart-Tablet. Provided to those with known severe respiratory / cardiac conditions or with early Covid-19 symptoms*
- ***Smart-Tablet device** with internet router if home WiFi not available. Locked down & remote support enabled.*

Figure 1.0.1: The Howz pathway for patient monitoring and potential benefits taken from TIHM report (KSS et al., 2021).

sion of 32% and 42% in care home admission (KSS et al., 2021). Also participants experienced a reduction in depression and anxiety. This shows the data collected from the device and the decisions made within the system can benefit both the user and the wider healthcare community.

Due to the increasing knowledge about dementia, what to monitor and focus on through the system of devices is clear. However, outside this dementia study, in a user-base of hundreds within a wider population, with differing behaviours, illnesses and circumstances, identifying change in well-being is more complex. Given the users may have differing needs, a method to monitor and identify changes in behaviour applicable to many users, needs to have a broader focus.

Developing methods on monitoring devices within health management, is a growing area of literature (Tun et al., 2021). One area of analysis includes the monitoring of vital signs using medical devices e.g., methodology that can accurately measure various vital signs (Khan et al., 2016). In practice, this would be equipment placed within the home that the household or carers would need to use to get measurements. Therefore, identification of changes relies on use of the devices and further reporting measurements above or below a certain threshold defined by medical professionals. Furthermore, the reported measurements from devices that flag an alert do not consider what information other sensors can provide. For example; a heart monitor, we may see a sudden jump in heart rate but other sensors show a lot of movement sensors triggered. Given the context this may not be cause for concern, as the jump in heart rate could be due to exercise.

When considering data on human activity, it is common to think of wearable devices, worn 24/7, that take several observations per second. Wearables have the advantage of high time resolution tri-axial data for the movement and location of an individual. Wearable technology can also monitor vital signs such as heart rate and act as a pedometer. This can automate the detection of changes in vital signs, as constant measurements can be taken and the known thresholds can be used to flag worrying changes in measurements. With the large amount of data, identification of activities performed by the user can occur (Qi et al., 2018) e.g., walking/sitting, eating and sleeping. A common method to identify activity is classification (Preece et al., 2009). One downside of classification methods, as the focus is classification of various activities, are they do not consider how a household can change over time.

For example, the activities may change over time and change in frequency. Overall these methods can provide interesting insight into the user, if we have access to huge datasets such as these. Wearable technology relies on the users wearing these devices 24/7 which can often lead to users feeling as if their privacy is invaded. Further, Kekade et al. (2018) show that adoption to wearables in the older population can be low.

An under utilised dataset within the network of devices are the static sensors. Placing passive infra-red motion, door and appliance sensors throughout the house has many benefits. They allow for monitoring of the households behaviour, how they move through the house and interact with sensors. Once set up the household does not have to think or worry about them as they get on with their day, allowing for a non-invasive way to monitor the household. This provides a time series of categorical observations that can provide information about a households' behaviour. These interactions with sensors produce binary (active/not active) data for each sensor.

There exists research into categorical datasets on an activity level for the Howz dataset (Taylor et al., 2021), however this does not consider information gained by considering sensors individually. Activity from different sensors, can provide different information; e.g. movement sensors - whether or not the household is sedentary versus appliance sensors - whether the household is eating meals. Reducing this to activity level diminishes the potential information we could gain.

Despite the potential benefits of using categorical sensor datasets to monitor behaviour, this data provides new challenges. There is little research on identifying changes within categorical sensor datasets for health management of the elderly and vulnerable population. However, this is not a problem restricted to Health Management. Pattern mining techniques have been used to identify changes in categorical sequences (Fournier-Viger et al., 2017). These methods, are predominantly used to identify interesting categorical patterns seen frequently in the data and not identify when these patterns change. Classification algorithms have been used to identify how similar categorical sequences are with applications in text data (e.g. spelling mistakes), DNA and proteins, and cyber security logs (Xing et al., 2010). These methods are designed for long sequences, generally struggling with smaller sequences that we could see within the passive sensor datasets. Changepoint analysis has been used to identify changes in time series data, however the main body of research relates to continuous data. There exists some research into categorical changes in time series (Höhle, 2010), the focus here is on identifying the probability each category will trigger, not considering a change in the event order (in context the sensor activations). Overall, new ways to identify change and understand behaviour utilising categorical sensor datasets over time is an open problem with applications in the health management and beyond.

In this thesis, we focus on broad methods functional over any household using the system of passive static sensors. We develop novel methodology to identify changes in passive sensor datasets, motivated by our industrial partnership with Howz. One of the aims for Howz in this setting, is personalised inference to the user so we focus on methodology that can highlight information to the household, carer or family member on the well-being of the household. The rest of this chapter introduces the reader to samples from the Howz data and performs exploratory analysis on different households, to provide motivation for the contributions to the thesis in Chapters 2-4.

1.1 Exploring the Howz Data

Our focus within the Howz product are the static infra-red motion sensors, door sensors and smart plug adapters with labels as to their location and purpose. Examples of these sensors include motion sensors; Landing and Lounge, appliance sensors; Toaster and Kettle, and door sensors; Front door and Fridge door. This data is considerably more sparse in comparison to wearables, as the user must interact with these sensors for information to be gained. The benefit for the user being, after set-up, these sensors are completely passive.

In this section we present 10 households with 9 weeks of data (1st of January to 3rd March 2020) provided by Howz, showcasing different identified behaviours with a range of sensors. The basic Howz package has a motion sensor, door sensor and smart plug, with the option to add more sensors. It is the choice of the user as to where they place the sensors within the house or if they place all of them. Howz advises the households to place the sensors on regularly used devices/locations. The households presented here have between 2 and 10 sensors within the house and are single occupancy.

Howz have previously considered the data on an activity level, performing analysis on low activity and predicting the getting up and going to bed for a household (Taylor et al., 2021). To take this further, our focus is sensor-specific, looking at ways to detect changes using the different sensors seen within the data. The sensor observations in the data are binary indications of presence or absence of activity for each sensor with a time stamp for the occurrence of each observation. We start by considering this to be categorical data with the observation of a sensor being the occurrence of that category.

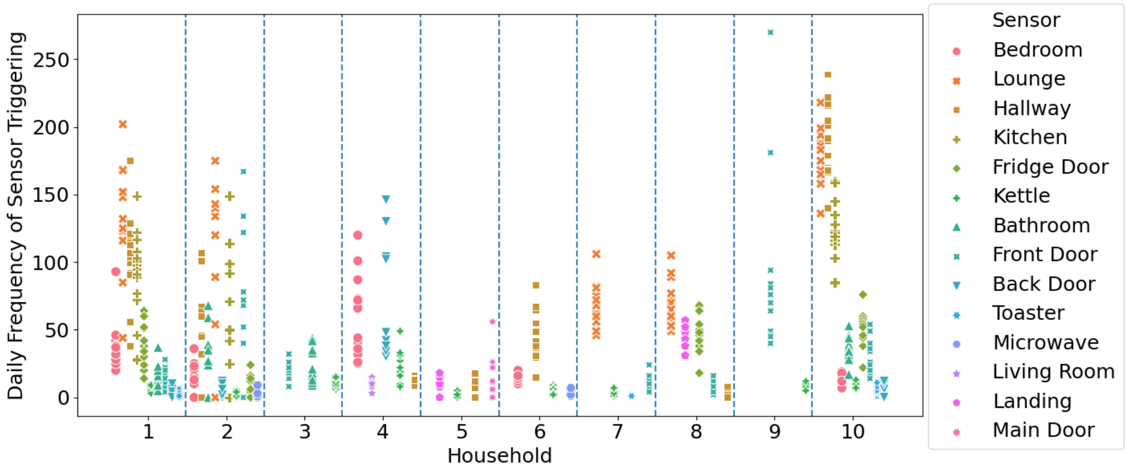


Figure 1.1.1: Ten households from Howz with 2 - 10 sensors with the daily frequency each sensor triggered for the 1st to 14th January (colour-coded).

Figure 1.1.1 shows the wide difference in the daily count for sensors triggering in different households. For example, Households 1, 2 and 10 have 10 sensors, with the motion sensors, Lounge, Kitchen and Hallway being triggered often over 100 times a day. Whereas, Household 3 and 5 have motion sensors, Bathroom and Landing, with fewer than 50 observations per day. This is a combination of the activity levels and the choice of sensors by the household, as some will be in areas frequently moved through.

There is a large difference in the number of observations seen in a day across households, seen in Figure 1.1.2, with some households averaging less than 50 obser-

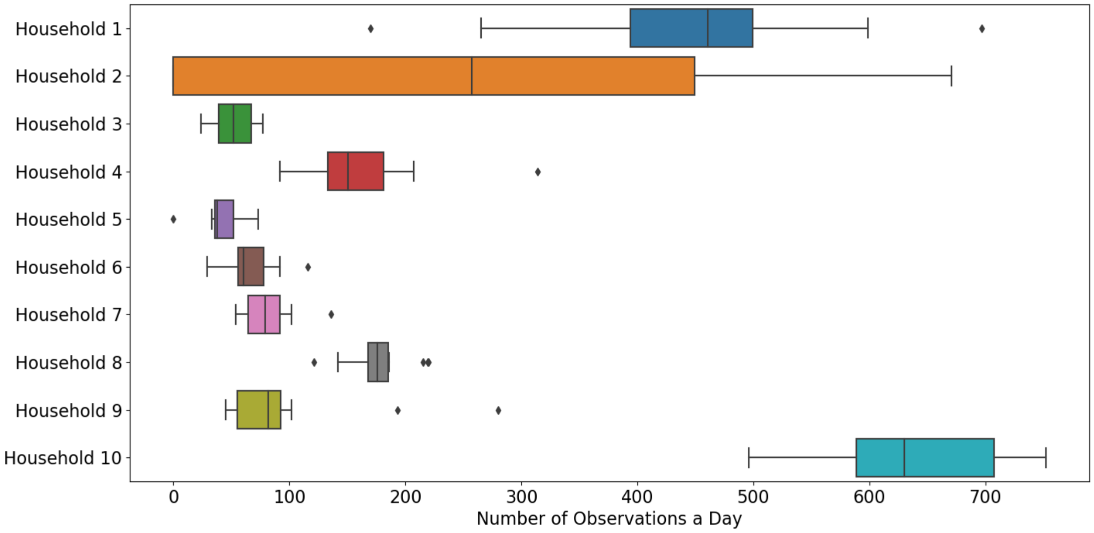


Figure 1.1.2: Box plot of the number of observations for each household from Howz.

vations per day. Therefore, we chose in Chapter 2 to focus on identifying changes in days of data, considering households with a small and varied number of observations each day.

Thus far we have looked at the number of observations throughout the day, without looking at the order of events within the day. Figure 1.1.3 shows Household 6, where we can see this household often gets up in the middle of the night, as well as having a clear wake up time, between 6 and 8am, and going to bed at around 10pm. The household seems to be very active up until lunchtime, with less observations seen in the evening and afternoon.

Having a daily routine is a well-known human behaviour and using routines to help understand and identify changes in behaviour seen within sensor data has had some success (van Kasteren et al., 2017). While, households are not likely to behave in the same way every day, small routines of behaviour are common due to how sensors are

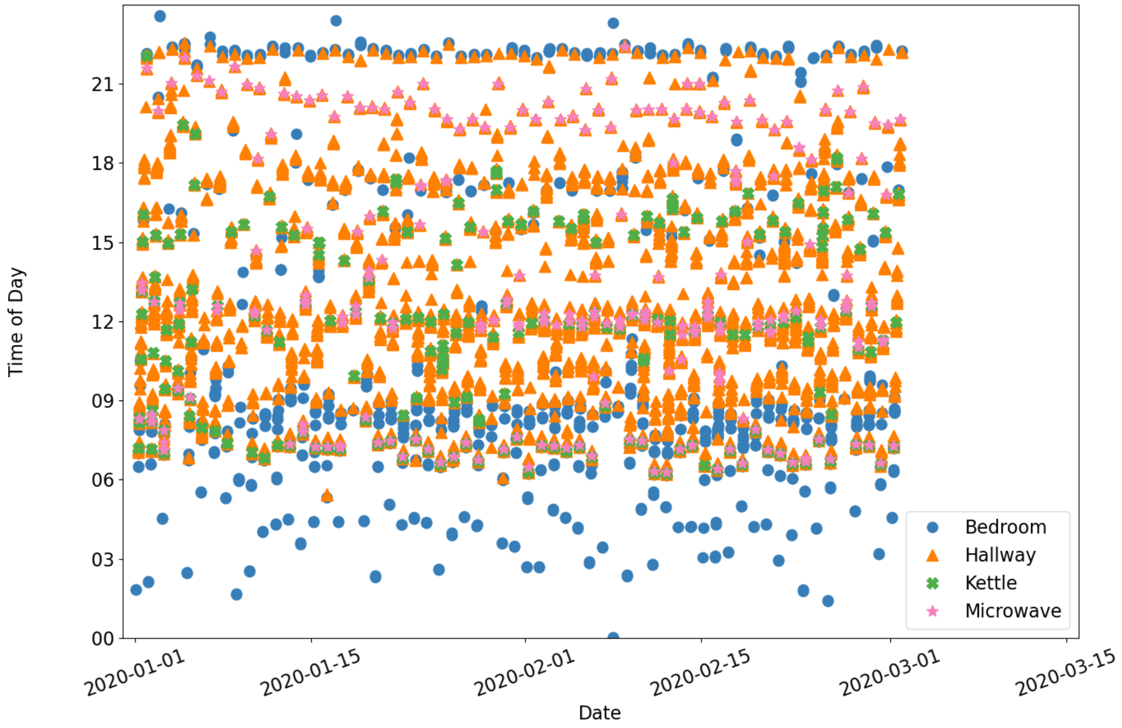


Figure 1.1.3: Colour-coded sensor events for Household 6 with 4 sensors; Bedroom, Hallway, Kettle and Microwave.

placed throughout the house and the minutia of daily life; getting up in the morning, eating meals, going to sleep. These small sequences of observations give a good representation of a household, and it is a cause for alarm when these sequences no longer occur. For Household 6, there are examples of these small routines throughout the day. In Figure 1.1.3 we can see the Hallway and Bedroom triggered as they go to bed. Figure 1.1.4 shows a zoomed in version of this household between 6 and 8 am. It is clear that they often go from the Bedroom, through the Hallway and use the Kettle (circled in purple). This is often followed by Hallway observations and the use of the Microwave. If a household was to perform fewer of these small routines

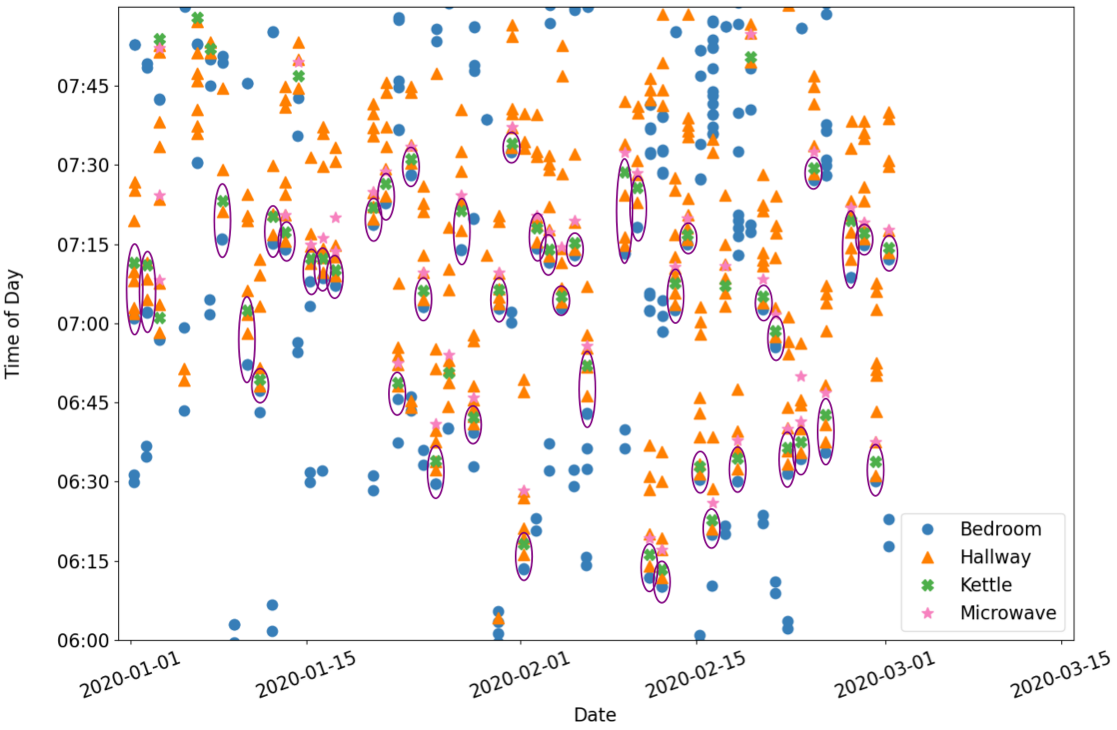


Figure 1.1.4: Colour-coded sensor events for Household 6 with 4 sensors; Bedroom, Hallway, Kettle and Microwave, between 6 and 8am. Purple circle to indicate small routine; (Bedroom, Hallway, Kettle, Hallway).

throughout the day, this could be an indicator of a change in behaviour within the household. This provides the following research objective for Chapter 2; to develop a novel method for identifying changes within sequences. Due to the nature of the data as discussed, the new methodology must consider:

- The natural routines seen within the sequences.
- Sequences of short and varied length.
- The varied behaviour seen day to day.

Looking at yes/no activation at the sampling level, the data would be heavily weighted to the 'no' category, as we would be looking at rare events. Hence, Howz often groups the data into 15 minute intervals throughout the day for analysis. If we wish to look at the behaviour of a sensor throughout the day, we can follow suit. We now consider the data as multiple binary time series, one for each sensor within the household. Taking the first month of data (January) for Household 6, we can look at the data on a 15 minute scale. We refactor the data such that a sensor is said to have triggered, if it is seen to trigger at least once in the 15 minute interval. Then we take the empirical probability (the frequency of the sensor over the total number of intervals). Figure 1.1.5 shows that the Bedroom sensor is likely to be triggered in the morning followed by a spike of Hallway activity. The Kettle and Microwave have clear spikes in probability in the early morning, lunchtime and evening. It is clear that there is a connection between these sensors, confirmed by the small routines seen in the data in Figure 1.1.4. Further to this, if we look at Figure 1.1.6, we can see that not only are there small routines within days, there also exist routines across days and weeks for certain sensors. In Figure 1.1.6, the Bedroom sensor is triggered in similar parts of the day across each week. Showing more clearly, that this household triggers the Bedroom door often when getting up in the morning, has some activity in the afternoon and a clear bedtime.

Considering the routines and the interactions between the sensors, the research objective for Chapter 3 is to design a novel method to model and forecast each sensor. This gives the following research aims for Chapter 3:

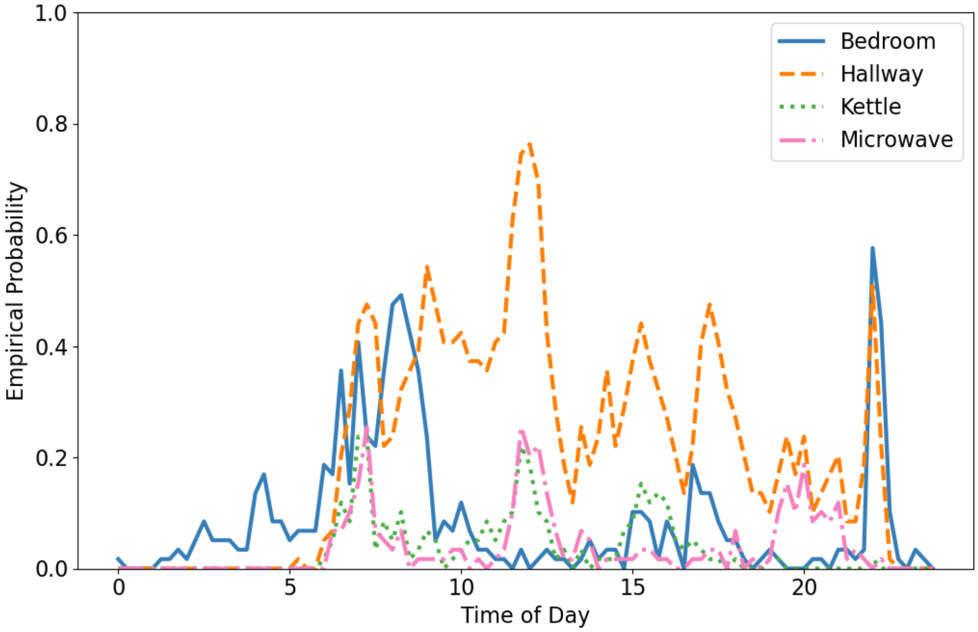


Figure 1.1.5: Empirical probability of the sensors for Household 6, triggering at each time interval in a day; Kettle (red), Microwave (pink), Bedroom (green) and Hallway (yellow) for a month of data.

- Consider the data on a sensor specific scale.
- Incorporate the time each sensor is triggered throughout the day.
- Account for data on previous days and from other sensors within the home.

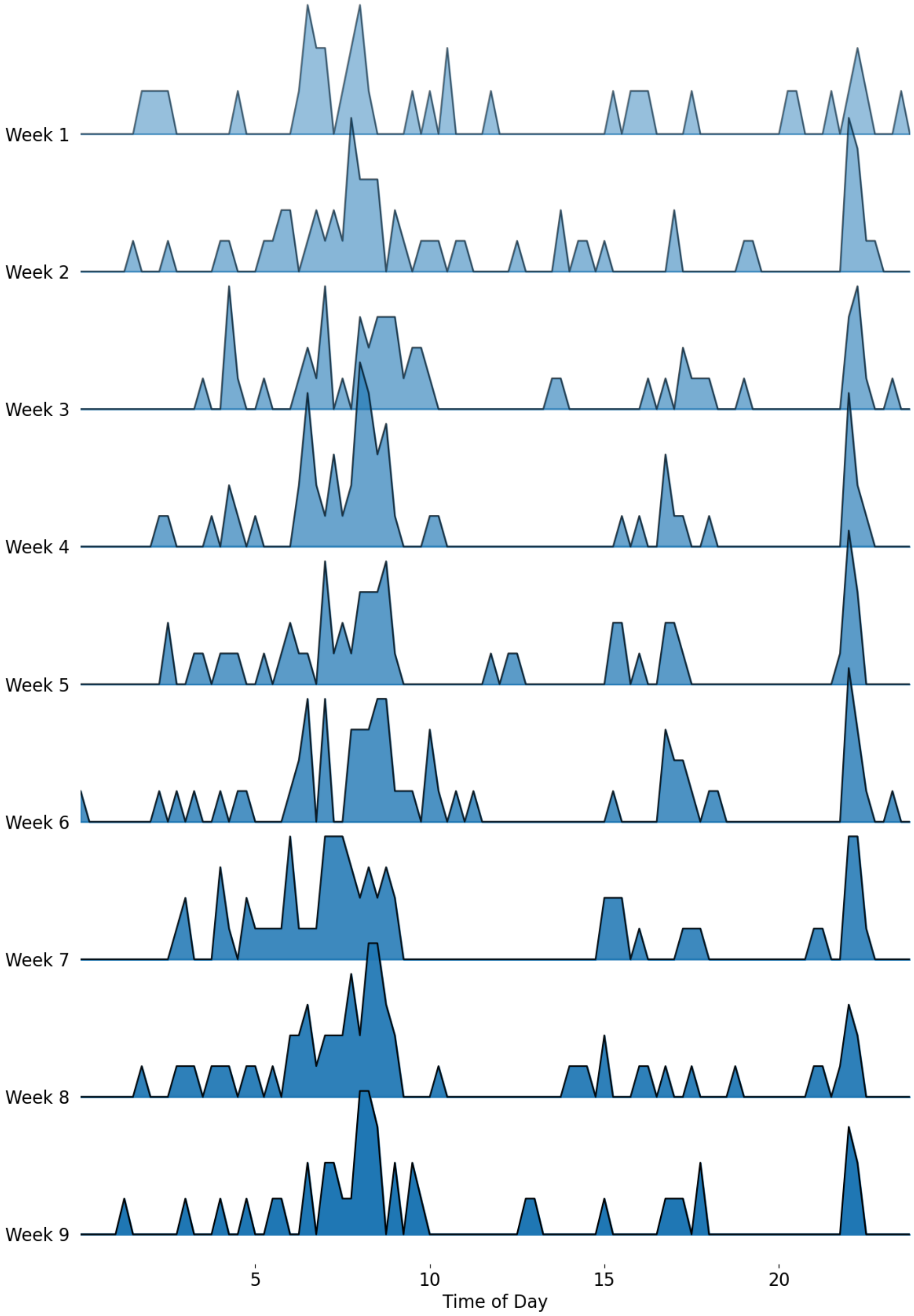


Figure 1.1.6: Empirical probability of the Bedroom sensor for Household 6, triggering at each time interval in a day over each week of data; Week 1 - January 1st-7th, Week 2 - 8th-14th etc.

Howz have a vastly increasing user-base and are interested in collecting similar households within a group structure. For example, sensors placed within care homes where each individual has a certain number of sensors assigned to them. Placing users into certain groups and monitoring if they deviate, can have many benefits in alerting the user and carers. For example; if a user triggered the Hallway sensor often throughout the day but changed over time to only being active in the morning, this could indicate the user tiring more quickly than before. In Figure 1.1.7, we can see the empirical Kettle usage across different households. For example, it may be of benefit to Howz to place households with constant behaviour across the day (Household 9 and 10) and high activity throughout the morning and afternoon (3 and 4) into different groups. This allows Howz to monitor if the behaviour changes, but this proves challenging with many different groups across hundreds of households. For example, with similar groups of households, i.e. morning/lunch/afternoon activity vs morning/lunch/evening activity, a household may be on the boundary between these two groups, appearing in either regularly.

To tackle this, we build upon Chapter 3 developing the research objective for Chapter 4; to provide a novel framework to monitor the probabilistic group membership of these households. The research aims for Chapter 4 are thus:

- Using known groups, identify how a household is related to each group.
- Identify when households deviate from those they were once similar to.
- Changes should not be identified when, e.g. a household lies on the boundary between two groups.

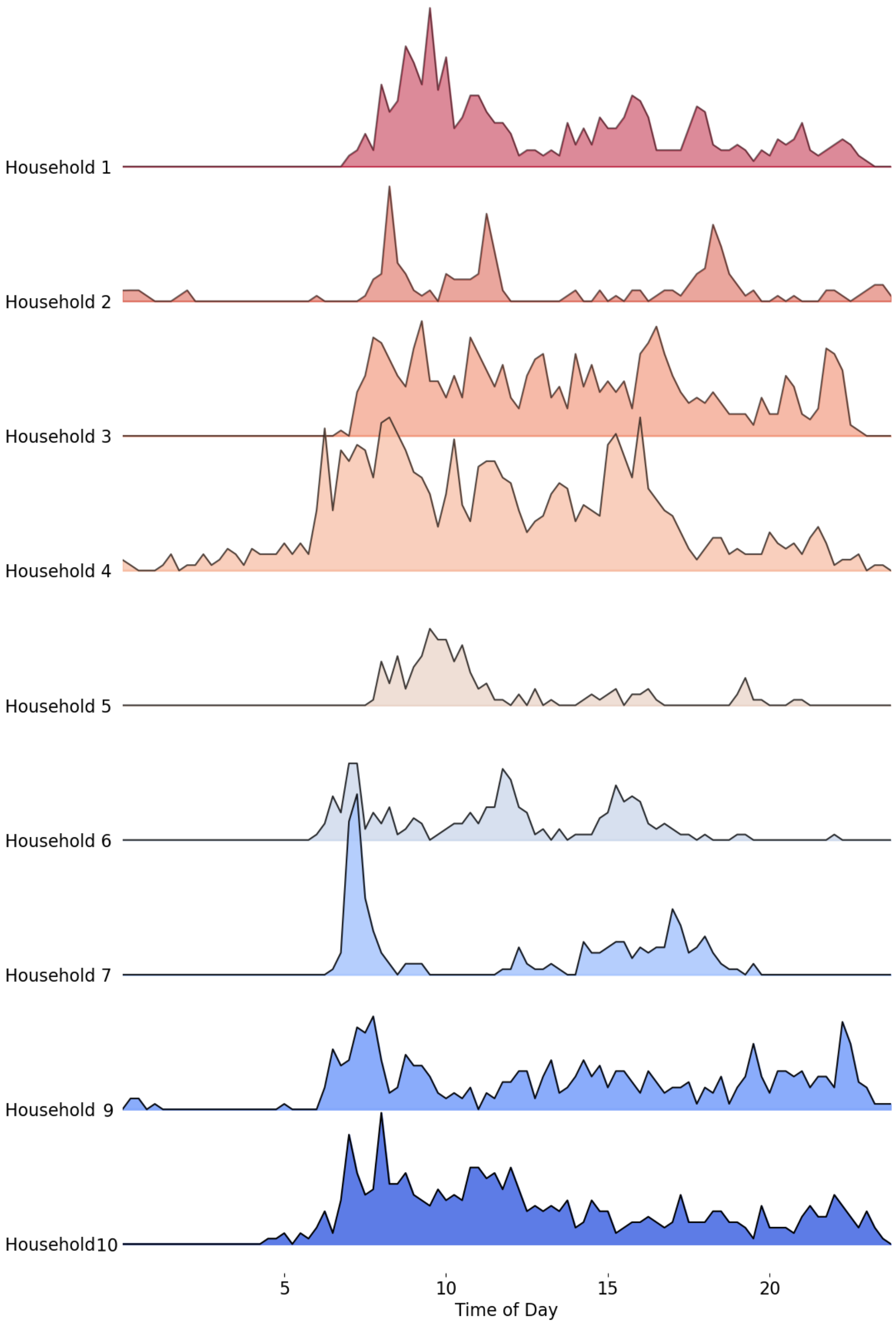


Figure 1.1.7: Empirical probability of the Kettle sensor triggering at each time interval in a day over the two months of data for each household. Note that Household 8 does not have a Kettle sensor.

1.2 Outline

This thesis first contributes a novel irregular sequence detection method in Chapter 2, which proposes a method for detecting subtle changes in sequences whilst taking account of the natural day-to-day variability and differing numbers of observations per day. In Chapter 3, we propose a novel method to model the probability a sensor will trigger throughout the day for a household whilst accounting for the prior data and other sensors within the home. The final contribution in Chapter 4, designs a framework to identify subtle changes in probabilistic cluster membership of households through time, combining clustering and changepoint methodology. Chapter 5 concludes the thesis with a summary of the contributions and discussion of further work to develop the methods described.

Chapter 2

Identifying Irregular Activity Sequences: an Application to Passive Household Monitoring

2.1 Introduction

Approximately 1 in 5 people will live to see their 100^{th} birthday due to advancements in modern medicine and other factors (Age UK, 2019). While this is a great achievement, it puts pressure upon the NHS and health care services due to the care that is necessary with an ageing population. As of April 2019, there are just under 12 million people in the UK aged 65 or over, with close to a quarter of those living by themselves. This group constitute 42% of elective admissions and 43% of emergency admissions to hospital (Age UK, 2019). It is clear that there is a growing need to help care for the elderly population as such pressures on the NHS and health care services will only increase in the future.

Increasingly, people are turning to technology to help improve health and care of the elderly. Howz is a product who uses unobtrusive passive sensors to ascertain patterns of behaviour and seeks to alert customers to help them to understand changes in their behaviours. Taylor et al. (2021) investigate changes in raw activity levels of Howz customers but do not consider the more detailed information collected about what activity is taking place. The larger changes in a person's routine can often be captured by changes in activity levels, but the more subtle changes in a loss of confidence or independence in living can be harder to discern. For example, a customer may routinely prepare hot lunch using the cooker but then switches to cold lunches as they become less confident in cooking. The activity levels will be the same but the nature of the task has changed and gives an indication of a decline in independence.

To identify these subtle changes we propose to consider a person's routine as their sequence of activated sensors per day and we wish to identify subtle changes in these routines. A significant challenge here is that a person's daily behaviour may have considerable variation and a varying number of activated sensors each day. Any method developed needs to be able to address these challenges.

2.1.1 Howz Dataset

When considering activity data it is common to think of a wearable device that samples at 100-500 observations per second, e.g. heart monitors, where the type of activity needs to be identified statistically. This relies upon the users wearing such devices 24-7. In contrast, the Howz data arise from static infra-red motion sensors, door sensors and smart plug adapters. This data is sparse in comparison to wearable devices, but the sensor is labelled with a location. Examples of labelled sensors include motion sensors; Landing and Lounge, appliance sensors; Toaster and Fridge door, and door sensors; Front door and Bedroom.

Figure 2.1.1 presents a household provided by Howz with three sensors installed. The data are binary indications of presence or absence of activity for each sensor with a time stamp for the occurrence of each categorical observation, taken from a sensor. The households provided by Howz for this research are single occupancy. Howz advises the households to place the sensors on regularly used devices/locations.

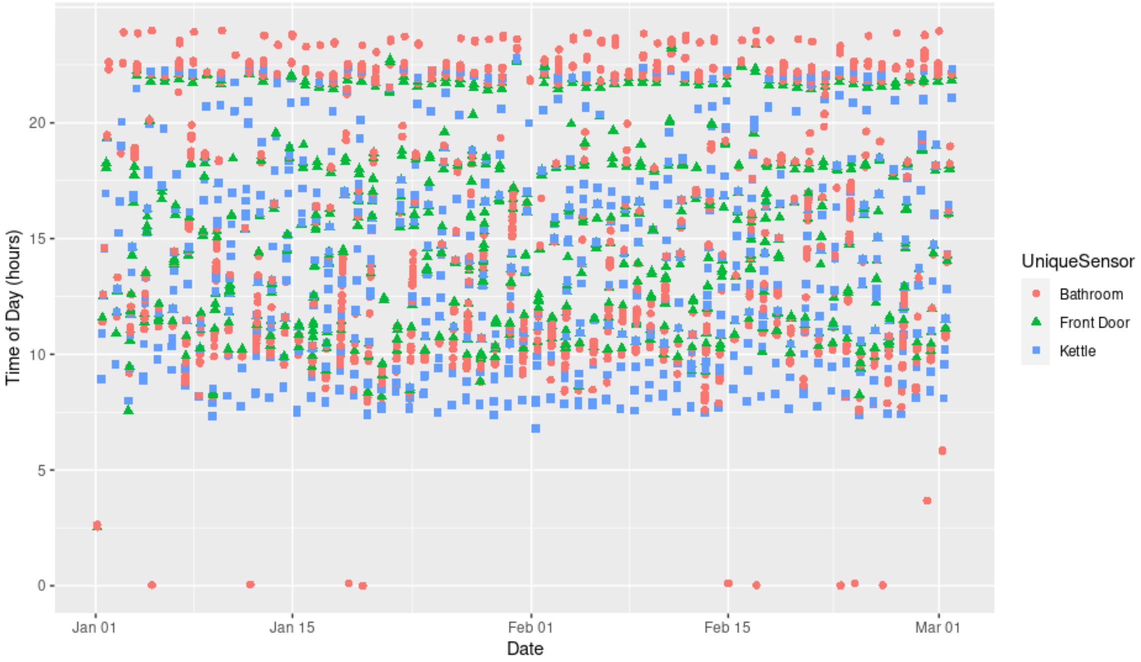


Figure 2.1.1: Colour-coded sensor events for a Howz household with 3 sensors; Bathroom, Front Door and Kettle.

In order to identify changes in sensor activation order day to day, we consider the nature of this data as a categorical sequence of events. We take each day to be a sequence of data, this will enable us to identify changes while considering the obvious daily structures, e.g., from Figure 2.1.1 multiple triggering of the Kettle followed by the Bathroom followed by the Front door. We define the common behaviour seen across days as a routine. An example of a regular routine would be waking up in the morning, putting on the kettle, letting the cat out and having some toast. This small routine could result in the following sequence: Hallway motion, Kettle, Back Door, Toaster. Many customers have very flexible routines, therefore, any method developed must be tolerant of alterations in this varied but regular behaviour. We aim to identify if changes in these routines occur.

2.1.2 Motivation and Structure

A common way to detect changes in behaviour is to use changepoint detection. Höhle (2010) presents methods to detect changes in categorical time series using multi-categorical regression models where they monitor the category probabilities seen within the data. Plasse and Adams (2019) propose to detect changepoints in multiple categorical data streams that monitor the category probabilities of a multinomial distribution. Generally these approaches focus on estimating changes in the probabilities of sensors activating within a single day but will not detect changes in the order of the events.

To focus on the ordering of events, He et al. (2019) and Fradkin and Mörchen (2015) derive methods for mining sequential patterns; identifying interesting subsequences and their frequency within the data. However, when deciding which subsequences are important, these methods do not consider that sequences can be very varied and still be part of the same underlying behaviour. Applying this method could potentially focus on sub-sequences which are important for characterising the broad routine but are not sensitive to identifying irregular behaviour.

Dafé et al. (2015) introduce a method to identify how similar two categorical sequences subject to noise are. The method allows for sequences to not match completely and still be classed as very similar by avoiding penalising too heavily on the mismatches within. This is ideal for our application, with the stipulation that the noise we need to account for is not meaningless. Dafé et al. (2015) are motivated by very large genomic sequences of the same length. In contrast, our sequences are relatively short and a household will be unlikely to have the same number and order of events each day. Figure 3.1.1 shows there are a different number of Bathroom and Kettle observations in the morning across January and February.

In this paper, we tackle the challenge of identifying changes within short categorical sequences with varied lengths and differing behaviour. We approach this problem by letting each day be a sequence of events for a household. To decide if a change in behaviour has occurred, we need a notion of typical behaviour for a household. Motivated by practices in Howz, we define this as a set of regular sequences for the household where each sequence would be a day of categorical sensor events. Our contribution is in developing a new metric for assessing the similarity of two categorical sequences of potentially different lengths. Then using this similarity measure to assess whether a new day is conforming to regular behaviour or should be flagged as irregular.

We develop a novel approach that directly addresses the challenges of a small number of sensors, small and varied sequence lengths, and variation within regular sequences to detect a change in behaviour. In section 2.2, we describe the set up for this problem and the proposed method is developed in section 2.3. Section 2.4 provides a simulation study and section 2.5 gives an example using the Howz dataset. In section 2.6, we summarise our results and discuss potential avenues for further research.

2.2 Problem Set Up

To define a change in routine for a household we must first consider what is regular behaviour. The method we propose assumes we have existing data to characterise regular behaviour for a household. We take multiple weeks of data, resulting in a set of regular sequences for every household, one per day.

Formally, let R denote the set of regular sequences that form the null distribution to compare against, and T be the set of new sequences to be tested. Let $X^R \in R$ be a sequence from the regular sequence set whose length is $|X^R|$. Furthermore, let $\mathcal{L}$ denote the library set of sensors such that X^R is defined as a sequence of categorical events:

$$X^R = \{x_j \ : \ x_j \in \mathcal{L} \text{ for } j = 1, \dots, \left|X^R\right|.\}$$

The notation similarly extends to the set of test sequences T. Reference to the regular and test sets are not essential when defining the sequence silhouettes (Section 2.2.1) and the similarity metric (Section 2.3). As such, explicit reference in the notation to the sequence sets is suppressed and only re-introduced where necessary. The aim is to take the vector sequences in T and assess whether, individually, they are consistent with regular behaviour (as described by set R). Typically the test set, T, will be of size 1.

In the following sections we use a similarity measure as a method for assessing if a sequence of activities is regular behaviour. However, there are a few application specific constraints on the data we must consider. Firstly, we have a small number of sequences in the set of regular sequences. This is because we want to keep the set of regular sequences up to date with the regular behaviour of the household. For example, if a person has a fall then their regular behaviour may change while they recover. If we are using, for instance, three months of regular behaviour then several new days may appear irregular due to the known change in behaviour. Once a change has been confirmed by the client we can use the new data as the regular set. Thus we use a smaller set of regular sequences that evolves over time, although how these evolve will not be addressed in this paper. We also need to acknowledge the small lengths of each sequence which vary day to day e.g., between 7 and 42 from the data in Figure 2.1.1. An example of why we may see such variation is a household which spends some days indoors and some outdoors. Hence, it is also important to acknowledge some households have very varied sequence patterns.

We present an approach for comparing two sequences that addresses these challenges in Section 2.2.1 before discussing how to calculate a similarity score based on these comparisons in Section 2.3.

2.2.1 Sequence Silhouettes

In contrast to numerical data, the similarity between categorical data sequences is not as trivial to define. Dafé et al. (2015) introduce a way to calculate the similarity by comparing subsequences within both the new sequence and regular sequence. This identifies the small routines (the subsequences) and bases the similarity score on if they appear within both sequences. Consider the example sequence X with $|X| = 7$ in Table 2.2.1, which consists of a library of 3 sensors: Door (D), Kettle (K) and Motion (M).

Index	1	2	3	4	5	6	7
Sequence: X	D	M	K	D	D	D	D

Table 2.2.1: Example of an activity sequence with Door (D), Kettle (K) and Motion (M) sensors.

First we must find all possible subsequences of X whilst preserving order. Clearly the number of subsequences increases exponentially when considering longer sequence lengths. As the number of observations per day varies considerably, we choose to impose an upper limit, $\mathcal{K} \leq |X|$, on the lengths of the subsequences to consider. The choice of $\mathcal{K}$ is examined in Appendix A.5.

Despite this upper bound, evaluating all possible subsequences of length $k \in 1, \ldots, \mathcal{K}$ scales poorly. Specifically, if the library set consists of $r = |\mathcal{L}|$ sensors, then there are up to $\frac{r(r^{\mathcal{K}}-1)}{r-1}$ possible subsequences to consider. To address this issue, Dafé et al. (2015) reduced the search space by using 'sequence silhouettes' that focus on the start and end of each subsequence for each possible subsequence length k. For example, the sequences DDK and DKK would be classed as belonging to the same sequence silhouette D*K.

Only counting unique silhouettes, for all $k > 1$ we would have a maximum of r^2 options and for $k = 1$ a maximum of r silhouettes. Therefore the total number of possible sequence silhouettes, P, in each sequence is $r + (\mathcal{K} - 1)r^2$. We give Definition 2.2.1 for each sequence silhouette (SQS), using a general sequence, X, for the condensed search method.

Definition 2.2.1. *(SQS)*

A sequence silhouette is defined by the triple $S = (s, s', k')$ *where items* $s, s' \in \mathcal{L}$ *respectively define the first and last categorical event of the subsequence that has length* $k = k' + 1$. *We define the set of all possible silhouettes as* $\mathcal{P} = \{S = (s, s', k') : s, s' \in \mathcal{L}, k' = 0, \ldots, \mathcal{K} - 1\}$. *The set of sequence silhouettes that exist for a given sequence,* X *with cardinality* $|X|$, *is defined as* $\mathcal{P}_X \subseteq \mathcal{P}$ *where* $S \in \mathcal{P}_X$ *if there exists some initial position* h *such that* $x_h = s$ *and* $x_{h+k'} = s'$. *It is possible that a sequence silhouette can match multiple times for a given sequence, and so we record the initial positions of all matches in the set* $\mathcal{H}_{X,S}$.

Returning to the example in Table 2.2.1, if we take X, an example of the sequence silhouettes found if we search through the sequence is presented in Table 2.2.2.

In Section 2.3 we propose a similarity metric that compares a single sequence within T and compares it to all the sequences within R. We will score each silhouette from a sequence separately and combine across sequences to assess if the occurrence

D	M	K	D	D	D	D	s	s'	k'
$\underline{\text{D}}$	M	K	D	D	D	D	D	D	0
D	$\underline{\text{M}}$	K	D	D	D	D	M	M	0
D	M	$\underline{\text{K}}$	D	D	D	D	K	K	0
⋮			⋮			⋮	⋮	⋮	⋮
$\underline{\text{D}}$	$\underline{\text{M}}$	K	D	D	D	D	D	M	1
D	$\underline{\text{M}}$	$\underline{\text{K}}$	D	D	D	D	M	K	1
D	M	$\underline{\text{K}}$	$\underline{\text{D}}$	D	D	D	K	D	1
⋮			⋮			⋮	⋮	⋮	⋮
$\underline{\text{D}}$	M	$\underline{\text{K}}$	D	D	D	D	D	K	2
D	$\underline{\text{M}}$	K	$\underline{\text{D}}$	D	D	D	M	D	2
D	M	$\underline{\text{K}}$	D	$\underline{\text{D}}$	D	D	K	D	2
⋮			⋮			⋮	⋮	⋮	⋮

Table 2.2.2: Sequence silhouette graphic for sequence X in Table 2.2.1; the left part of the table gives the sequence with the start and end of each silhouette (underlined) and the right part of the table gives the sequence silhouette triples (s, s', k') as defined in Definition SQS.

of a silhouette in the new sequence is statistically different from those in R.

2.3 Proposed New Similarity Metric

Using the sequence silhouette framework, we design a similarity score to decide how similar two sequences are by looking at the sequence silhouettes they share and the information inside them. Table 2.3.1 gives an example of a sequence silhouette two sequences, X and Y, have in common.

Index	1	2	3	4	5	6	7	8	SQS		
Sequence: X	M	$\underline{\text{D}}$	M	K	D	D	D	$\underline{\text{D}}$	D	D	6
Sequence: Y	$\underline{\text{D}}$	K	M	D	M	D	$\underline{\text{D}}$	D	D	D	6

Table 2.3.1: Example of two sequences with sequence silhouette $(D, D, 6)$ with the start and end of the sequence silhouette underlined.

When making a comparison between sequences we consider all instances of a sequence silhouette. However, to increase computational efficiency we first created an indexing system which can decide whether a sequence silhouette is in the sequence.

In order to assess how similar the two sequences are we consider two factors, the presence of the sequence silhouette, and the elements within it. We take inspiration from Dafé et al. (2015) who designed a score function which penalised heavily on internal mismatches. However, for household behaviour small differences in the internal ordering are not as important (Figure 2.1.1) and so we do not expect strict routines to exist. Therefore, we want to penalise a mismatch between the two subsequences in a different way than in Dafé et al..

In Section 2.3.1 we design a new similarity score for each silhouette. Section 2.3.2

then proposes an adjusted score which takes into account the relative sizes of the sequences. This requires the maximum and expected scores which are not trivial to calculate and are considered in Section 2.3.4 and 2.3.5 respectively.

2.3.1 Similarity Score Function

The metric for the similarity score for comparing one instance of a sequence silhouette match between two sequences is given by Definition 2.3.1.

Definition 2.3.1. *(Similarity Score)*

Consider two categorical sequences X *and* Y*. Let* $k \in \{1, \ldots, \mathcal{K}\}$ *be the length of the subsequence the sequence silhouette,* $S = (s, s', k')$*, refers to, with* $k' = k - 1$*. For all shared sequence silhouettes* $S \in \mathcal{P}_{X,Y} = \mathcal{P}_X \cap \mathcal{P}_Y$*, define* $\mathcal{H}_{X,Y,S} = \mathcal{H}_{X,S} \times \mathcal{H}_{Y,S}$ *to be the set of all paired initial positions* $(h, g) \in \mathcal{H}_{X,Y,S}$ *so that* $x_h = y_g = s$ *and* $x_{h+k'} = y_{g+k'} = s'$*.*

Given scaling factors, $\beta \in \mathbb{R}_{>0}$ *and* $\lambda \in \mathbb{R}_{>0}$*, we define the similarity metric as,*

$$\Gamma(X \mid Y, h, g, S) = \beta + \begin{cases} \sum_{c=1}^{k'-1} \gamma(x_{h+c}, y_{g+c}) & \text{if } k' > 1, \\ 0 & \text{otherwise.} \end{cases}$$

where γ *is the recursive function:*

$$\gamma(x_{h+c}, y_{g+c}) = \begin{cases} \gamma(x_{h+c-1}, y_{g+c-1}) + \lambda & \text{if } c \geq 1 \text{ and } x_{h+c} = y_{g+c}, \\ \gamma(x_{h+c-1}, y_{g+c-1}) & \text{otherwise.} \end{cases}$$

We give an example in Table 2.3.2 to show how the inside of the sequence silhouette (SQS) is calculated. This shows that the score would be $\beta + 4\lambda$ for the silhouette

Aligned Index	1	2	3	4	5	6	7	SQS		
Sequence: X	$\underline{\text{D}}$	M	K	D	D	D	$\underline{\text{D}}$	D	D	6
Sequence: Y	$\underline{\text{D}}$	K	M	D	M	D	$\underline{\text{D}}$	D	D	6
Score		0	0	λ	λ	2λ				

Table 2.3.2: Evaluation of the similarity score between two sequences based on the sequence silhouette $(D, D, 6)$, resulting in a score of $\Gamma(X \mid Y, h, g, (D, D, 6)) = \beta + 4\lambda$.

match in this example. Definition 2.3.1 gives the similarity score for one match of a silhouette. Across all matches the similarity measure is:

$$\begin{aligned}\Gamma(X \mid Y, S) &= \sum_{(h,g)\in\mathcal{H}_{X,Y,S}} \Gamma(X \mid Y, h, g, S) \qquad (2.3.1)\\ &= |\mathcal{H}_{X,S}|\ |\mathcal{H}_{Y,S}|\ \beta + \begin{cases} \sum_{(h,g)\in\mathcal{H}_{X,Y,S}} \sum_{c=0}^{k'-1} \gamma(x_{h+c}, y_{g+c}) & \text{if } k' > 1,\\ 0 & \text{otherwise.}\end{cases}\end{aligned}$$

Equation 2.3.1 is the similarity score for one silhouette for all instances between two sequences X and Y. By re-introducing reference to the set of regular sequences, the total similarity of sequence X^R against all other sequences within R with respect to a given silhouette is:

$$\Gamma(X^R \mid R, S) = \sum_{Y\in R\setminus X^R} \Gamma(X^R \mid Y, S). \qquad (2.3.2)$$

The similarity metric for a test sequence from T, X^T, is likewise defined by comparing it against all sequences in R:

$$\Gamma(X^T \mid R, S) = \sum_{Y\in R} \Gamma(X^T \mid Y, S). \qquad (2.3.3)$$

2.3.2 Adjusted Similarity Score

Intuitively, due to the large variation in the sequence length of Howz data, large sequences are likely to have more sequence silhouettes and thus, have larger similarity scores when compared to shorter sequences. Therefore, we need to introduce a scaling method to mitigate the effect of varying sequence size. Taking inspiration from the Rand Index used in clustering (Rand, 1971; Hubert and Arabie, 1985), we correct for the larger variation in sequence length using the expected and maximum scores as follows.

Definition 2.3.2. *(ASC) The Adjusted Similarity Score for a sequence X and sequence silhouette S:*

$$ASC(X \mid S) = \frac{\textit{Similarity Score - Expected Score}}{\textit{Maximum Score - Expected Score}} = \frac{\Gamma(X \mid S) - \mathbb{E}(\Gamma(X \mid S))}{\max(\Gamma(X \mid S)) - \mathbb{E}(\Gamma(X \mid S))}$$

In Section 2.3.1, we defined the similarity score. In the following sections we explore the maximum and expected scores.

2.3.3 Maximal Packing

To calculate a maximum score, we must first consider the maximum number of times a silhouette could be present in a sequence, X of length $|X|$, the maximal packing. A first naive attempt at an upper bound would be $|X| - k'$ as we consider the overlapping of sequence silhouettes, i.e. a sequence silhouette instance may start within another instance. For example, using X defined in Table 2.3.1, there exists two silhouettes

$(D, D, 2)$ which overlap between index 4 and 7. However, there are ways to improve upon this maximum bound. When considering types of overlap, there are two different types of silhouette: a silhouette where the start and end are the same ($s = s'$) and one where they are different ($s \neq s'$). For the first case, it is clear the maximum number of silhouettes occurs when the sequence is entirely the same event. The second case is less intuitive, we give the definition for the maximal packing here with the full derivation in appendix A.1.

Proposition 2.3.3. *(Maximal Packing) For the different cases, the maximum number of occurrences for a sequence silhouette, S, within a sequence, X, where $|X|$ is the length of the sequence,*

$$\textit{Maximal Packing} = \begin{cases} |X| - k' & k' \geq 0 \quad s = s', \\ k' \left\lfloor \frac{|X|}{2k'} \right\rfloor + w_X & k' \geq 1 \quad s \neq s'. \end{cases}$$

where $\lfloor n \rfloor$ denotes the integer part of the number n. The term w_X describes how many additional matches of the SQS after perfect overlapping. Specifically, the residual part of the sequence after accounting for perfect overlapping has length $v = |X| \mod 2k'$, where the modulo operator returns the remainder of the Euclidean division. Hence:

$$w_X = \begin{cases} 0 & v < k \\ v \bmod k' & \textit{otherwise.} \end{cases}$$

An example of maximal packing for the silhouette $(D, K, 2)$ for a sequence of length 7 would be $DDKKDDK$ with a count of 3 occurrences of the silhouette: two in one instance of perfect overlapping, and one occurrence in the remaining sequence.

2.3.4 Maximum Score

Given the number of sequence silhouettes gained from the maximal packing, we next need to compute the associated maximum score. For the first case, we assume that each sequence silhouette instance within the maximally packed sequence will have a match at each element. However, this does not hold true for the second case (see Table A.2.2, Appendix A.2). The overlap results in a lower score. This is because the score for each silhouette match in the $2k'$ space will get smaller, due to increased mismatches within the sequence silhouette. The following definition describes the maximum score for these cases. The derivation is given in the appendix A.2.

Theorem 2.3.4. *(Maximum Score) The maximum overall score for all instances of the sequence silhouette, S, in the two sequences, X, Y, with cardinality $|X|$ and $|Y|$, respectively. Therefore, the maximum score* $(\max(\Gamma(X \mid Y, S))$ *when comparing these sequences is:*

$$\max(\Gamma(X \mid Y, S)) = \begin{cases} \beta\ |X|\ |Y| & \textit{if } k' = 0, \\ (|X| - k')\ (|Y| - k')\left(k'\beta + \frac{k'(k'-1)\lambda}{2}\right) & \textit{if } k' \geq 1 \textit{ and } s = s', \\ \beta k \lfloor \frac{|X|}{2k'} \rfloor k' \lfloor \frac{|Y|}{2k'} \rfloor & \textit{if } k' = 1 \textit{ and } s \neq s', \\ \left(M^{\star} \left\lfloor \frac{|Y|}{2k'} \right\rfloor + \sum_{c=1}^{w_Y} M(c)\right)\left(k' \left\lfloor \frac{|X|}{2k'} \right\rfloor + w_X\right) + A(w_X, w_Y) & \textit{if } k' > 1 \textit{ and } s \neq s', \end{cases}$$

where $A(w_X, w_Y)$ is defined in Appendix A.3 to account for subsequent score matching. We define the function M by:

$$M(c) = \begin{cases} \beta + \frac{\lambda(k'-c-1)(k'+c)}{2} & k' > 1,\ 0 \leq c < k' - 1, \\ \beta & k' > 1,\ c = k' - 1, \end{cases}$$

where c *has maximum value* w_Y*. For ease of notation we denote* $M^\star$*:*

$$M^\star = k'\beta + \begin{cases} \lambda & \textit{if } k' = 2, \\ \frac{\lambda k'(k'-1)(k'+1)}{3} & \textit{if } k' > 2, \\ 0 & \textit{otherwise.} \end{cases}$$

We now have a method to calculate the maximum score between two sequences, and hence, the maximum score over all the regular sequences is:

$$\max(\Gamma(X \mid S)) = \sum_{Y \in G} \max(\Gamma(X \mid Y, S)).$$

To ensure the maximal score derivation holds we add the following constraint. For a fixed $\mathcal{K}$*,*

$$\sum_{c=1}^{\mathcal{K}-1} \gamma(x_{h+c}, y_{g+c}) < \beta,$$

such that β *and* λ *do not change across sequence silhouettes.*

2.3.5 Approximated Expected Score

For each sequence silhouette, when trying to calculate the expected score for the new sequence using the set of regular sequences, we quickly run into computational challenges. The search space would increase greatly with k, as the inside of the silhouette must be considered. As minimizing computational complexity is a key priority we choose to approximate the expected score. A reasonable approximation would need to take into account the relative probability of the silhouette whilst scaling with the number of sensors. Thus we propose estimating the expected score as follows.

Definition 2.3.5. *(Approximated Expected Score) Considering sequence $X \in R$, and sequence silhouette $S = (s, s', k)$, the approximated expected score is defined as:*

$$ES(X \mid S) \; = \frac{r^2 P(s) P(s')}{|R|} \times \; \max(\Gamma(X \mid S)),$$

where r is the size of the library $\mathcal{L}$ and $P(\cdot)$ is the estimated probability of seeing a certain event using the regular behaviour sequences, i.e. the relative frequency of events. This can be defined similarly for a sequence within set T.

2.3.6 Final Algorithm

Using the approach described in Sections 2.3.1-2.3.5, we calculate the similarity scores for a new sequence, X, within T comparing it to the regular set. This produces one similarity score for each sequence silhouette in a sequence. In order to assess whether these scores are significantly different from the regular set we need to compare the set of regular sequences to each other. We take a sequence within the regular set, R, and compare it to the remaining sequences in the set using the method in Section 2.3. Then repeat for all the regular sequences. Note, we divide by the total number of sequences being compared for the new sequence and regular sequences, due to the extra comparison for the new score.

We estimate the null distribution of each sequence silhouette score using those calculated from the regular set via a kernel density (Racine and MacKinnon, 2007; Davison and Hinkley, 1997). We use this null distribution to obtain a p-value for each silhouette, $S \in \mathcal{P}_X$. However, if we were to simply reject the new sequence if a p-value is below significance level α, typically the size of the test will not be maintained due to multiple testing. Instead, a multiple testing procedure with the following hypothesis test is used for testing sequence X:

$$
\begin{aligned}
H_{0,S} &: \text{Sequence silhouette, } S \in \mathcal{P}_X\text{, is regular for the household.}\\
H_{1,S} &: \text{Sequence silhouette, } S \in \mathcal{P}_X\text{, is irregular for the household.}
\end{aligned}
$$

The new sequence is labelled as irregular for our application if one of the hypotheses, $|\mathcal{P}_X|$, is rejected, after correcting for the multiple hypotheses. More emphasis should be placed on correctly finding all sequences with irregular behaviour, rather than rejecting a few regular behaviour sequences. Contextually, this is to ensure the safety and well-being of the household. Hence, a multiple testing procedure which controls the False Discovery Rate (FDR) is used. We choose a procedure from Benjamini and Hochberg (1995) which focuses on controlling the FDR at significance level α using a sequentially modified Bonferroni correction. We calculate the p-values for each sequence silhouette using the estimated null distribution, before ordering them in ascending order, assign indexing $v_{(q)}$ for $q = 1, \dots, |\mathcal{P}_X|$. Let,

$$
\hat{p} = \max\left(q \in \{1, \dots, |\mathcal{P}_X|\} \ : \ v_{(q)} \leq \frac{q\alpha}{|\mathcal{P}_X|} \right).
$$

If $\hat{p}$ does not exist we fail to reject any of the hypotheses and consider the new sequence as regular. Else, we reject $H_1, \dots, H_{\hat{p}}$ and label the sequence as irregular. A condensed pseudo-code for the Irregular Activity Sequence Identification algorithm is given in Algorithm 1.

Irregular Activity Sequence Identification

```
Input:       Regular set $R$ with sequences $X_i^R$ for $i = 1, \dots, |R|$.
             Test set $T$ with sequences $X_j^T$ for $j = 1, \dots, |T|$.
             Sequence silhouettes, $S \in \mathcal{P}$.
             Significance level $\alpha$.
             Parameters $\beta$ and $\lambda$ in evaluating similarity score function $\Gamma$.
Initialise:  $U = \emptyset$, $V = \emptyset$.

Iterate For $X_j^T \in T$:
  1. For each $S \in \mathcal{P}_{X_j^T}$:
     (a) Calculate the average similarity score, $\frac{\Gamma(X_j^T \mid R,S)}{|R|}$ based on equation 2.3.3.
     (b) Scale using $ASC(X_j^T \mid R, S)$.
     (c) For $i \in \{1, \dots, |R|\}$:
         i. Calculate the average similarity score, $\frac{\Gamma(X_i^R \mid R,S)}{|R|-1}$ based of equation 2.3.2,
            where sequence $X_i^R$ is excluded from $R$.
         ii. Scale using the $ASC(X_i^R \mid R, S)$ and add to set $U$.
     (d) Evaluate the kernel density estimate of $U$ to define the null distribution.
     (e) Calculate p-value using the density estimate and add to set $V$.
```

Algorithm 1 Irregular Activity Sequence Identification

```
Continued

  2. Order $V$ in ascending order and assign indexing $v_{(q)}$ for the $q^{th}$ smallest p-value.

  3. Calculate,

        $\hat{p} = \max\left(q \in \{1, \dots, |V|\} \ : \ v_{(q)} \leq \frac{q\alpha}{|V|}\right).$

     If $\hat{p}$ does not exist we fail to reject $X_j^T$ as a regular sequence.

     Otherwise reject the null hypothesis for sequence $j$ and label as irregular.

Output: $X_j^T$ labelled as regular or irregular.
```

2.4 Simulation Study

We now consider how Algorithm 1 performs on simulated data. The method is implemented in R, (R Core Team, 2020), for daily sequences under a range of scenarios, in order to test for sensitivity. We focus on the basic package offered by Howz i.e., a library consisting of 3 sensors. Let these sensors be Door (D), Kettle (K) and Motion (M) arbitrarily. We will focus on a variety of types of household starting with those with no visible routines i.e., independent data.

There are many components that we can vary to mimic the real data example in Section 2.5. We need to consider how to simulate the entries with different probabilities for each sensor, the varied lengths of sequences and how many sequences are in the regular set. For all simulations, we report significance level $\alpha = 0.05$.

2.4.1 Households with No Visible Routines

To simulate independent days, the data are generated based on the probabilities for each category. Due to the short sequences, a randomly generated sequence can exhibit wildly different probabilities than that which it was generated from. Thus rather than considering random sequences we generate a sequence such that the proportions approximately match the probabilities of the sensors being triggered and then generate random permutations of this sequence for the different days. For example, a sequence of length 10 with probabilities (0.8,0.1,0.1) would have a sequence MMMMMMMMDK which would be randomly permuted.

To keep the set of regular behaviour up to date and to reduce computation time, we consider having 8 weeks, 4 weeks and 2 weeks of regular sequences i.e. $|R|$. Let the upper bound of the length of SQS be $\mathcal{K} = 3$. Then, choose $\beta = 1$ and $\lambda = 0.5$ to ensure it fits the maximal packing constraint in Section 2.3.4. In simulating the data, we consider varying the sequence lengths by sampling the sequence lengths uniformly between a minimum $c \in 4, 10, 25$ and maximum $d \in \{10, 25, 50\}$. The minimum lengths are greater than or equal to the number of sensors to ensure that each sensor has a chance of being seen in a daily sequence. Let the length of each sequence, X, be $|X| \sim \text{Unif}(c, d)$ where c and d are the minimum and maximum length of the sequence respectively. The elements of the categorical sequence X, x_j for $j \in \{1, \ldots, |X|\}$ are Door, Kettle and Motion with the vector $\boldsymbol{\rho}$, of length 3, being the probabilities of triggering each of the 3 sensors independently at each j. Table 2.4.1 presents three potential independent households with 3 sensors. The vectors

Household	$\boldsymbol{\rho_1}$	$\boldsymbol{\rho_2}$
A	$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$	(0.5,0.4,0.1) , (0.8,0.1,0.1)
B	$(0.8, 0.1, 0.1)$	(0.7,0.15,0.15) , (0.6,0.2,0.2) , $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$
C	$(0.65, 0.25, 0.1)$	(0.8,0.1,0.1) , $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$

Table 2.4.1: Details of the household probability sets used in simulation.

$\boldsymbol{\rho_1}$ denote the probabilities for each sensor and $\boldsymbol{\rho_2}$ are the changes in routine we consider. The first set are small changes in probability, with the final one simulating a large change in behaviour. We also simulate new sequences using $\boldsymbol{\rho_1}$ to test the false detection rate.

We perform 500 replications for each set of parameters, simulating the sequence lengths and observed sensors separately each time. The values in Table 2.4.2 are the proportions of sequences labelled as irregular, calculated using algorithm 1.

Regular Sequence Probabilities	New Sequence Probabilities	$\lvert R\rvert$	Minimum Sequence Length 4	10	25
A: $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$	$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$	56	0.038	0.02	0.00
	$(0.5, 0.4, 0.1)$		0.918	1	1
	$(0.8, 0.1, 0.1)$		0.93	1	1
A: $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$	$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$	28	0.058	0.024	0.018
	$(0.5, 0.4, 0.1)$		0.918	1	1
	$(0.8, 0.1, 0.1)$		0.928	1	1
A: $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$	$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$	14	0.08	0.026	0.042
	$(0.5, 0.4, 0.1)$		0.918	1	1
	$(0.8, 0.1, 0.1)$		0.928	1	1
B: $(0.8, 0.1, 0.1)$	$(0.8, 0.1, 0.1)$	56	0.022	0.024	0.018
	$(0.7, 0.15, 0.15)$		0.236	0.812	1
	$(0.6, 0.2, 0.2)$		0.928	1	1
	$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$		1	1	1
B: $(0.8, 0.1, 0.1)$	$(0.8, 0.1, 0.1)$	28	0.062	0.036	0.020
	$(0.7, 0.15, 0.15)$		0.354	0.784	1
	$(0.6, 0.2, 0.2)$		0.93	1	1
	$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$		1	1	1

Table 2.4.2: Each column represents (from left to right); the probabilities the regular sequences are simulated with, the probabilities the new sequences are simulated with, number of regular sequences and the different minimum lengths of the sequences. The elements of the table contain the proportion which reject at least one hypothesis. Maximum sequence length 25.

Regular Sequence Probabilities	New Sequence Probabilities	$\lvert R \rvert$	Minimum Sequence Length 4	10	25
B: $(0.8, 0.1, 0.1)$	$(0.8, 0.1, 0.1)$	14	0.070	0.048	0.032
	$(0.7, 0.15, 0.15)$		0.428	0.758	1
	$(0.6, 0.2, 0.2)$		0.922	1	1
	$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$		1	1	1
C: $(0.65, 0.25, 0.1)$	$(0.65, 0.25, 0.1)$	56	0.036	0.036	0.042
	$(0.8, 0.1, 0.1)$		0.846	1	1
	$(1/3, 1/3, 1/3)$		1	1	1
C: $(0.65, 0.25, 0.1)$	$(0.65, 0.25, 0.1)$	28	0.070	0.070	0.054
	$(0.8, 0.1, 0.1)$		0.852	1	1
	$(1/3, 1/3, 1/3)$		1	1	1
C: $(0.65, 0.25, 0.1)$	$(0.65, 0.25, 0.1)$	14	0.082	0.062	0.05
	$(0.8, 0.1, 0.1)$		0.842	1	1
	$(1/3, 1/3, 1/3)$		1	1	1

Table 2.4.3: Each column represents (from left to right); the probabilities the regular sequences are simulated with, the probabilities the new sequences are simulated with, number of regular sequences and the different minimum lengths of the sequences. The elements of the table contain the proportion which reject at least one hypothesis. Maximum sequence length 25.

In general, from Tables 2.4.2 and 2.4.3, we see that the method has good power for this range of household behaviours. For Household B with the very small change in behaviour from (0.8,0.1,0.1) to (0.7,0.15,0.15), the method struggles to pick up the change when the minimum length of sequence is 4. This is most likely due to the small change in probability resulting in no difference in the simulated new sequence compared to regular sequences for the varying sequence lengths. The false positive rate is controlled as the number of regular sequences increases for these households. As expected, the power increases as the size of the change increases. The false positive rate also increases as the number of sequences in the regular set decreases. Given the small number of regular sequences, only a small number of scores are used to approximate the null distribution. This means we are likely to see new sequences of regular behaviour in the tails of the distribution. Therefore, this method is likely to have a higher false detection rate than the significance level when the regular set is small. We also varied the length of $\mathcal{K}$, β, λ and increased the number of sensors. These did not make a material difference to the results (see appendix A.5 for reference).

2.4.2 Example of a Household with Regular Routines

We consider a household with varying lengths of sequences and a regular routine. A regular routine refers to a household with a few small common routines present in most of the days in the set of regular sequences.

We simulated each regular sequence using the sequence in Table 2.4.4 as a template. The sequence has a few small routines, to give an example of regular behaviour. To simulate variations in this sequence, at each position with probability 0.4 the event

Position	1	2	3	4	4	6	7	8	9	10	11	12	13	14	15
Sequence	D	D	D	D	K	M	D	D	D	K	D	K	M	D	D
Position	16	17	18	19	20	21	22	23	24	25					
Sequence	D	K	K	D	D	D	M	K	D	D					

Table 2.4.4: Regular sequence template used in Section 2.4.2 and 2.4.3.

is removed. This results in a noticeable difference in sequences without changing the template sequence too drastically. To simulate regular behaviour for the user, we use the same underlying process for the new sequence as the regular sequences.

We consider three changes in behaviour for this household. The first and second change use the sequence in Table 2.4.4 as a template to simulate the regular sequences. At each position in the sequence there is a probability of $\frac{1}{3}$ that the event is removed and a probability of $\frac{1}{3}$ that it is replaced with an event simulated with equal probability. Alternatively, at each position in the sequence there is a probability of 0.2 that the event is removed and a probability of 0.5 that it is replaced with an event simulated with equal probability of each sensor. For someone with a regular routine, an example of a large change we would want to flag as irregular would be the loss of routine. We simulate this randomly by setting the probability of each sensor as equal, varying sequence lengths between 4 and 25 uniformly. We again perform 500 simulations for each set of parameters in the table.

The results in Table 2.4.5 show similar false positive rates to the households with no visible routine. The power of the method appears to be driven by the modification

New Sequence	$\lvert R \rvert$	Proportion of Sequences labelled Irregular
Simulated Like Regular Sequences	56	0.032
	28	0.044
	14	0.062
Simulated with Probability $\frac{1}{3}$ Removed and $\frac{1}{3}$ Replaced	56	0.65
	28	0.70
	14	0.73
Simulated with Probability 0.2 Removed and 0.5 Replaced	56	0.798
	28	0.832
	14	0.854
Randomly Simulated with Probabilities $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$	56	0.96
	28	0.972
	14	0.974

Table 2.4.5: Results for the flexible routine example.Each column represents (from left to right); how the new sequences are simulated, number of regular sequences and the proportion which reject at least one hypothesis.

of elements in a routine rather than the absence of elements which is our intention in the design of the similarity score. This is well aligned with the application to detecting changes in routines.

2.4.3 Example of a Household with Strict Routines

Consider a household with a strict routine, with little variation day to day. Using the same sequence in Table 2.4.4, we follow the same procedure as in Section 2.4.2. However, at each position in the sequence up to 24, with probability 0.3 it is swapped with the event ahead of it. This gives new sequences which are more similar to the initial sequence in Table 2.4.4 but reflect subtle changes in behaviour such as using the Kettle before the Toaster. We simulate the change in behaviour in the same way as Section 2.4.2. We perform 500 simulations for each set of parameters.

New Sequence	$\|R\|$	Proportion of Sequences labelled Irregular
Simulated Like Regular Sequences	56	0.048
	28	0.090
	14	0.112
Probability $\frac{1}{3}$ Removed and $\frac{1}{3}$ Replaced	56	0.886
	28	0.894
	14	0.868
Randomly Simulated with Probabilities $(1/3, 1/3, 1/3)$	56	1
	28	1
	14	1

Table 2.4.6: Results for strict routine example. Each column represents (from left to right); how the new sequences are simulated, number of regular sequences and the proportion which reject at least one hypothesis.

Note that due to the strict routines simulated there will be smaller variance in the regular set. This results in a slightly higher false positive rate than in previous examples - although it is still controlled. This also means that smaller deviations will be flagged as irregular behaviour and thus the power in this case is higher than in previous setting.

Overall the simulations demonstrate that the variability in the regular set is key to controlling the false positive rate. The larger number of sequences in the regular set also decreases the false positive rate but 28 sequences is viable. The larger the regular set, the more power we have to detect smaller changes in behaviour.

2.5 Examples using the Howz dataset

We show two households provided by Howz. The first household has 49 days of data with 3 sensors; Bathroom motion, Front Door and Kettle. The second household has 88 days of data with 4 sensors; Hallway motion, Kettle, Main door and Microwave. We take two weeks as the regular set as this is typical practice within Howz. Similar to the simulation study, we set $\mathcal{K} = 3$, significance level $\alpha = 0.05$ and choose $\beta = 1$ and $\lambda = 0.5$. To calculate the estimated probability of a sensor being seen for the expected score, we use the proportion seen in the regular set.

Figure 2.5.1 presents the sensor activity and the results for our first household. From a separate analysis on overall activity and the start/end of a users day, Howz has seen that there have been no discernible changes in the data. The method proposed here echoes these results, with no silhouettes flagged as irregular. This shows the

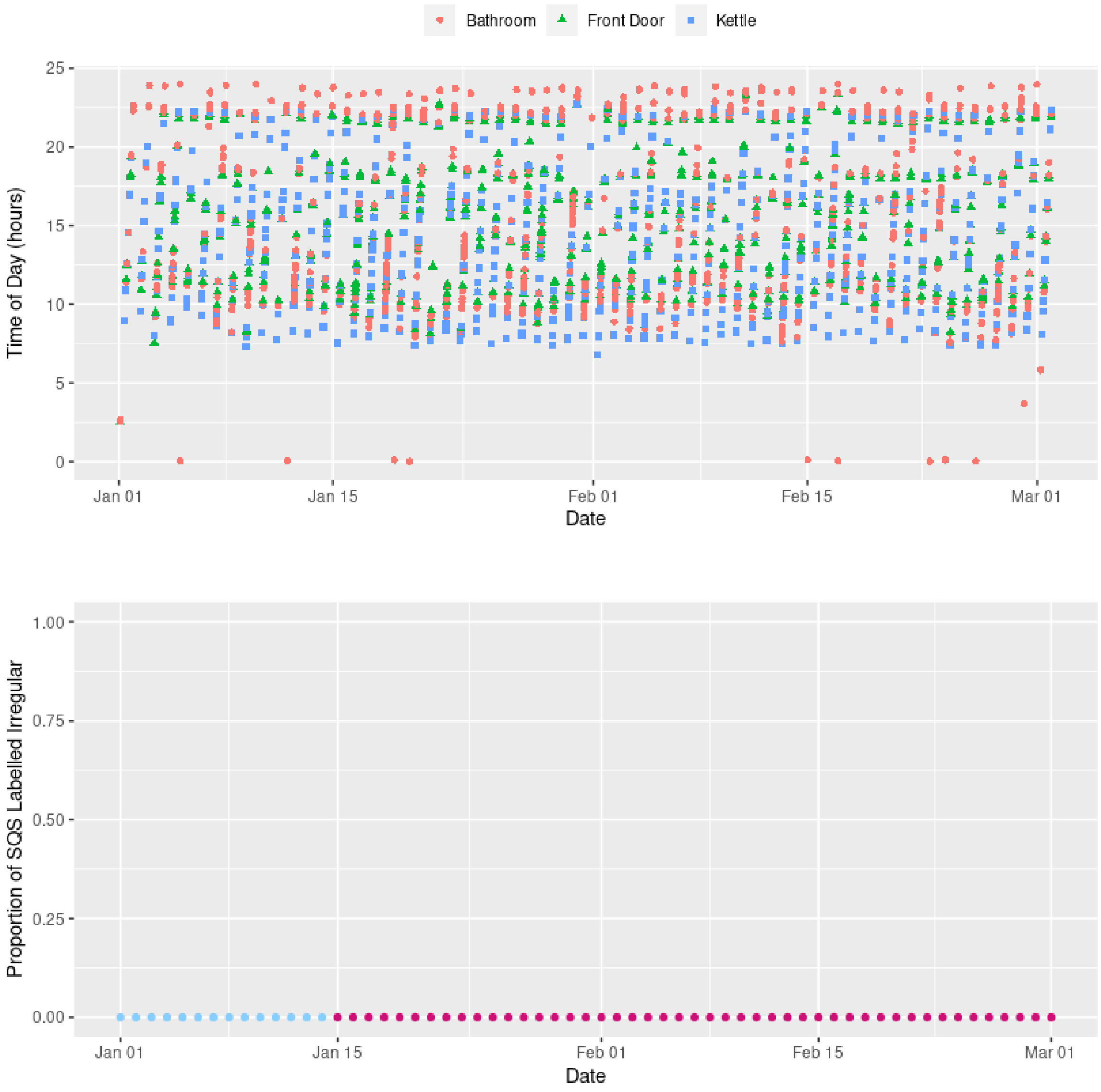


Figure 2.5.1: Top figure: Household with 3 sensors, showing the period between January and March. Bottom figure : Results with dots indicating regular set (light blue) and proportion of sequence silhouettes labelled irregular (red).

method, even with a small regular set, does not flag any changes for this household.

We report the computational time for this example, to show how fast this method performs on a household. We look at this in two parts; time taken to calculate the regular sequence similarity scores and the test for whether a new sequence is irregular or not. We expect in practice, the similarity scores for the regular sequences to be calculated offline so that we can use Irregular Activity Sequence Detection algorithm on the daily data as it comes in. Hence, the computational time for the regular sequence similarity scores for this example is approximately 112 seconds (virtual machine with 4 cores, 3Ghz processor and 16Gb RAM) and the average time to run the method on the new days i.e., the days between July 8^{th} and October 10^{th}, as around 9.9 seconds. Using simple parallelisation on the regular sequence similarity scores, using 3 cores we can reduce this to approximately 38 seconds. However, for Howz it may be more useful to run the method on households in batches simultaneously on independent cores, rather than within the method.

Figure 2.5.2 present results which Howz have prior knowledge of a change occurring in the household. Howz proposed this household as an example due to the fact that they had confirmed events in the middle of August and September with the household. A key feature of the events is a lack of any sensor activation, as the person was no longer living in the house. From Figure 2.5.2(top) we can clearly see a change in sensor behaviour prior to the first gap in sensor data. The household changes from activity throughout the day and night, to just morning and evening activity, dominated by the Main Door sensor. Looking closer, 3 or 4 days prior to this, we can see that the Main door occurs more frequently in the evening and the Kettle earlier in the morning. Whilst our analysis does not take into account the timing of the sensors, the bottom figure provides interesting additional insight. We pick up the event occurring in the middle of August with the proportion of irregular silhouettes increasing until intervention. The initial irregular silhouettes are highlighted a few

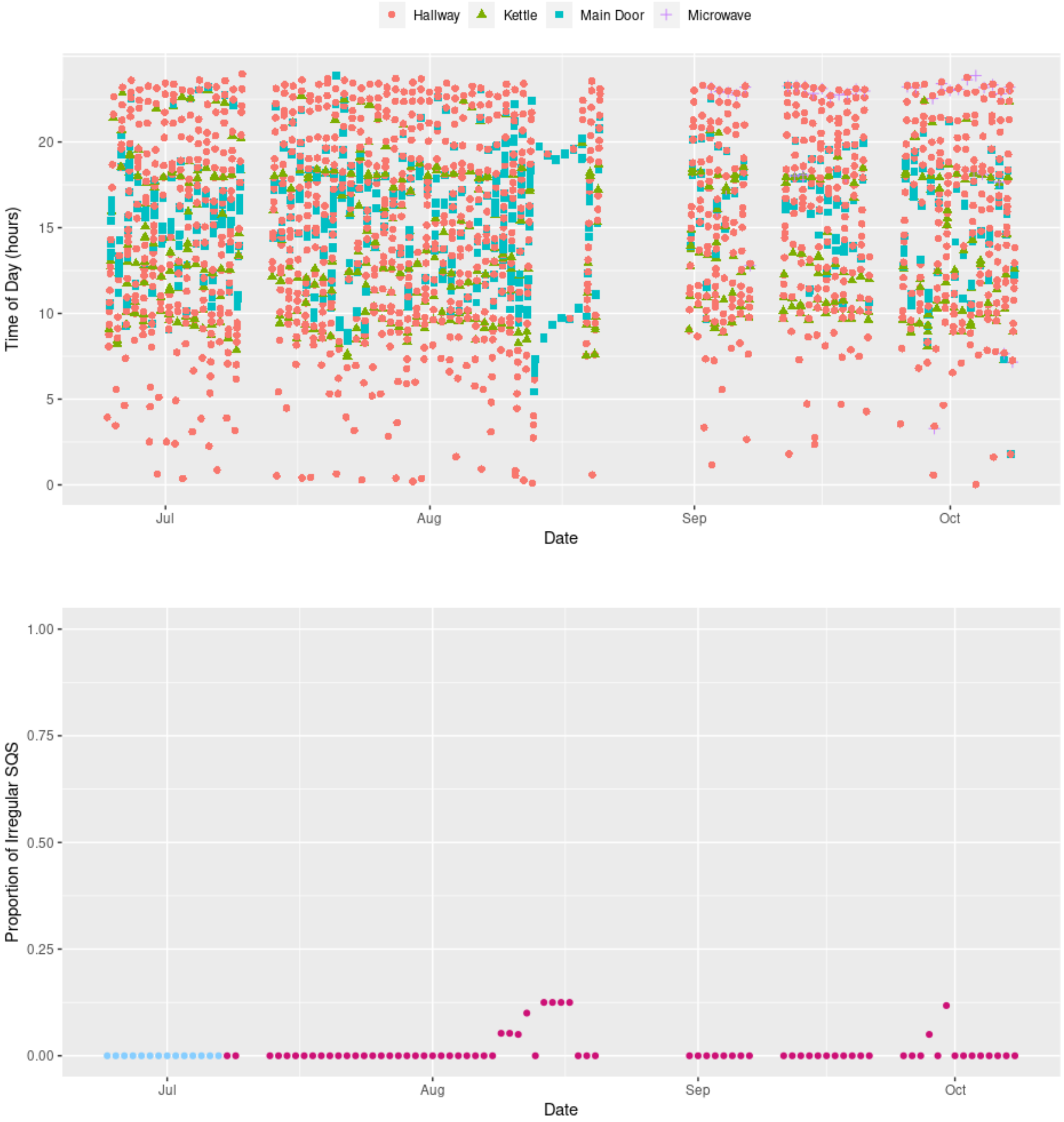


Figure 2.5.2: Top figure: Household with 4 sensors, showing the period between end of June and mid October. Bottom figure : Results with dots indicating regular set starting end of June (light blue) and proportion of sequence silhouettes labelled irregular (red) out of total possible for that sequence.

days prior to the clear morning/evening change which looks to capture the more subtle changes highlighted earlier.

Leading up to and during the event, the method flags the silhouettes (Hallway, Hallway, 0) and/or (Hallway, Kettle, 1) as irregular. The first indicates that overall the Hallway movement decreases meaning that the individual is less active in the Hallway relative to the sequence lengths. In conjunction with the decrease of Hallway, Kettle as a routine this could indicate a change in the activity level in the household. This was the conclusion reached by Howz when considering other analyses of aggregated data.

The silhouettes flagged as irregular towards the end of September are (Hallway, Hallway, 0) and then (Main Door, Main Door, 0), (Main Door, Main Door, 1) respectively. The first is due to an increase in Kettle and Main Door observations and as such the methodology is expecting an increase in Hallway sensors which does not manifest. The second is due to a decrease in the Main door sensors from a typical value of around 8-20 to just 2 observations in conjunction with Kettle and Hallway maintaining their usual frequency of activation.

This example shows the method is able to detect a change in routine before the change would usually be flagged by Howz alongside providing a more targeted description of what activity has changed. Naturally the description (Hallway, Hallway, 0) is not appropriate to send to a user and so some degree of interpretation of the change by the Howz team is required.

2.6 Conclusion

We have presented a novel method to detect changes in the behaviour of households on a day to day basis using sparse categorical data. The method is able to detect changes in new sequences that may be varied in length, including very small lengths. In practice for the Howz dataset, we cannot expect more than 8 weeks of regular behaviour. We have demonstrated that the method works with just two weeks of regular data, although increasing the number of regular days does improve the false detection rates, as expected. The model is very adaptable such that it can be applied to different situations with different constraints by changing the upper bound on the subsequence length, $\mathcal{K}$ and the partial matching parameters β and λ to suit the specific purpose and risk profile of the household. For example, if it was only important that the start and end of the silhouette occurred, λ could be set to 0, for applications where the inside of the silhouette holds no meaning. In practice, we would not update the regular sequences every day, which accounts for the majority of the computation time. After the offline calculation of the scores for the regular sequences, as discussed in Section 2.5, the method is fast enough that Howz could run hundreds of households overnight. We have shown this could be sped up further using parallelisation. However, within the industrial context, simultaneous batch processing of households on independent cores is likely preferred.

A more accurate representation of the expected score would account for the potential dependence between the sensors. However, the adopted definition was favoured for computational ease, which performs well in the simulation study. The households presented here are single occupancy, a future avenue of research would be considering multiple occupancy and how whether individual changes can be identified within the household data.

Currently the method is used on daily activity sequences but it is easily extended to weekly or monthly sequences. This could potentially provide insight into the longer term behaviour of the household. However, this will have computational drawbacks. Alternatively, we could further develop the approach to incorporate the time aspect explicitly into the model. One potential approach would be taking specific periods within the day and running the analysis on these subsequences to identify changes in behaviour. This would allow us to place more emphasis on distinguishing between two sequences e.g., DKMDDDDKMK where the initial KM routine was in the morning verses at lunchtime. Another approach would be including a time component to the similarity score. The method will still pick up a change due to the sequence silhouettes around this routine changing but timing changes may be identifiable.

Chapter 3

Modelling and Forecasting of at Home Activity in Older Adults using Passive Sensor Technology

3.1 Introduction

Life expectancy in the UK has increased since the 19th Century (ONS, 2018). As of 2019, there are just under 12 million people in the UK aged 65 or over (Age UK, 2019), with close to a quarter living by themselves. Those living on their own are more prone to trips to the Emergency Room and are more likely to suffer from mental illnesses (Dreyer et al., 2018). An ageing population such as this results in extra pressures for the NHS and health care services. By 2030 it is likely that one in five people will be aged 65 or over (ONS, 2017). It is clear there is a growing need for new ways to help the ageing population and reduce strain on the NHS and health care services.

3.1.1 Motivation

Howz is a product which introduces a number of non-intrusive sensors within the homes of older people and care facilities. The data received from these sensors is used to understand a household's behaviour, which can be used to improve the health and well-being of the household. For example if a customer gets up during the night frequently it could indicate a water infection. Once Howz has detected this, they make the user, family member or carer aware, so that action can be taken, reducing severity and preventing hospitalisation.

In this paper we focus on estimating the probability a sensor will trigger in a 15 minute interval given the other household sensors. We look at this information on a 24 hour scale, as previous analysis (Taylor et al., 2021) indicates a household is likely to follow similar routines throughout the day. For example; in the process of getting up we could expect the bedroom door, bathroom door and kettle sensor. Therefore, if the bedroom and bathroom sensors are triggered, we would expect an increase in the estimated probability of the kettle sensor.

Howz Anonymised Dataset

Figure 3.1.1, is an example of data for a household with 4 sensors; lounge, toaster, kettle and front door. Howz typically provide 3 sensors as a starter package for a new customer, with the option to install additional sensors. It is the customer who chooses how many sensors are used with some households having up to 10 sensors. Each data point has a sensor type and whether the sensor was triggered within the 15 minute interval.

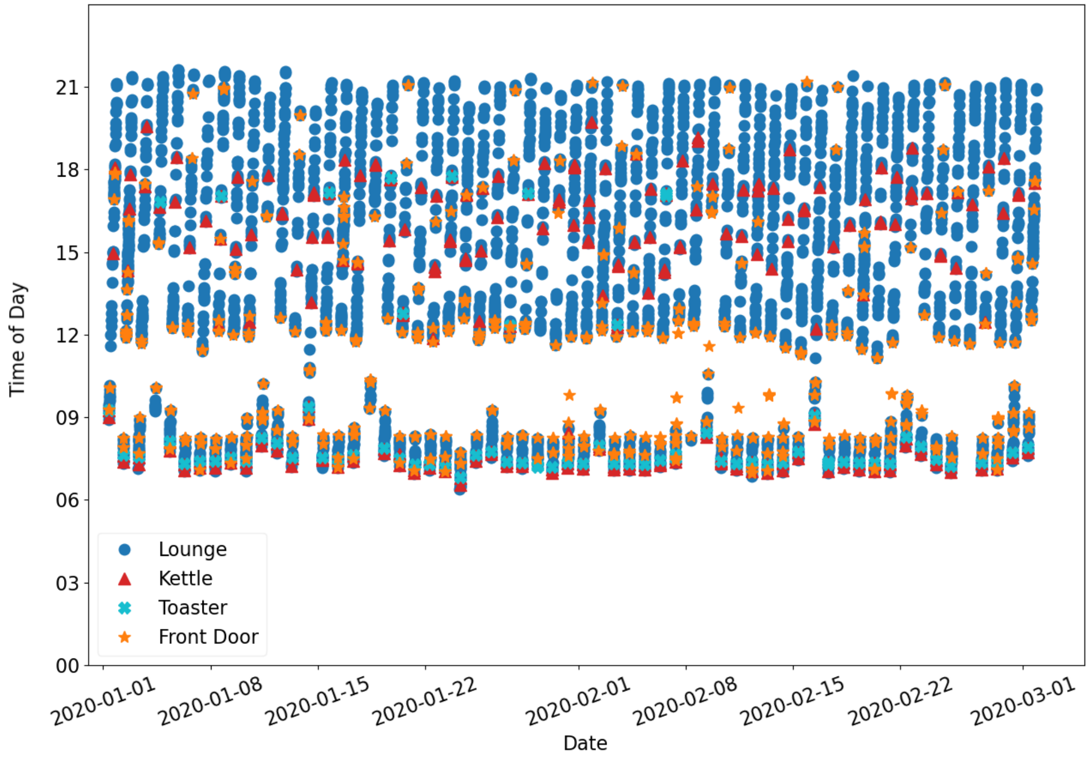


Figure 3.1.1: Colour-coded sensor events for a household; Lounge, Toaster, Kettle and Front Door.

We are interested in presenting the data on a 24 hour scale to explore how the household changes over the day. Figure 3.1.2 shows an example of this empirical probability in February for the kettle and toaster sensors. From the empirical estimate we can see that the probability varies throughout the day with the highest between 6am to 9am and the kettle having low level activity in the afternoon until 8pm with very little activity for the toaster after 9am.

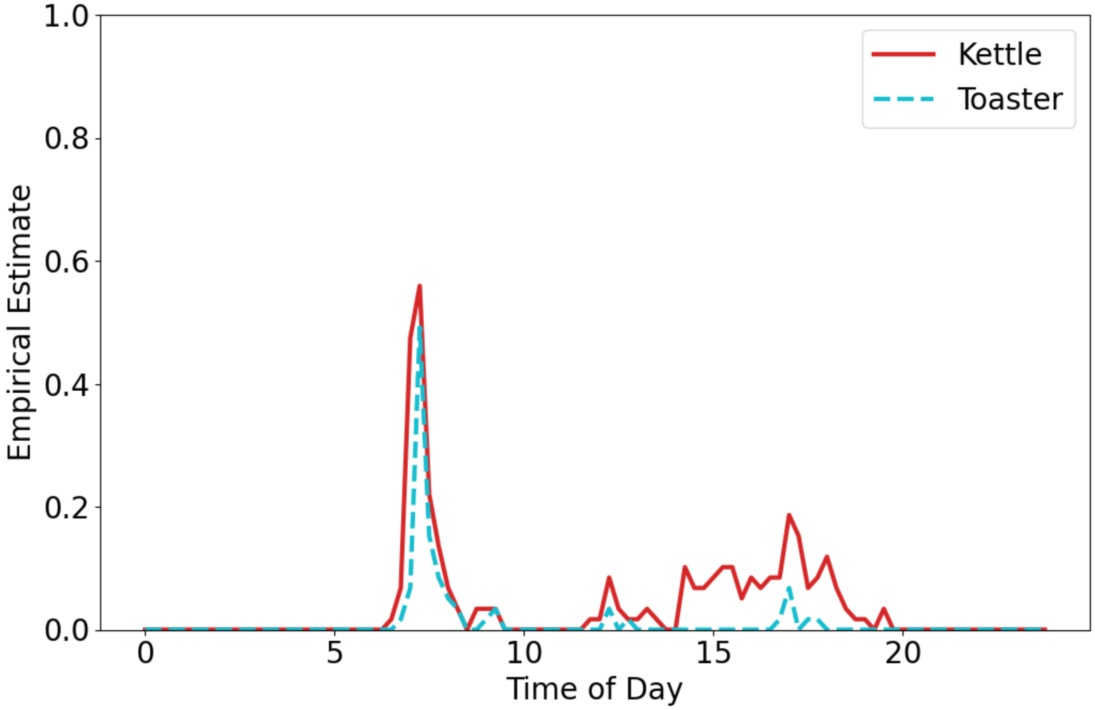


Figure 3.1.2: Empirical probability of the Kettle sensor triggering at each time interval (red line) vs Toaster sensor (blue).

3.1.2 Statistical Motivation and Structure

For discrete time series data, the main body of research concerns time series analysis of count data (Davis et al., 2021). See Handbook of Discrete-valued time series (Davis et al., 2016) for a general book on methods for discretised data. However, for our application it is necessary to look at the data within the day, rather than the daily count, and how that changes day to day. For example, using kitchen sensors at different times within the day could be a sign of forgetting to eat until the user becomes very hungry.

Households tend to follow small routines throughout the day, i.e. getting up and getting ready for the day and getting ready for bed. Hence, particular sensors are likely to be seen in sequence with each other. There has been recent research into classifying such routines with the use of wearable technology and other sensors with notable research papers gathered in Lotfi et al. (2020). In contrast, our focus is on incorporating potential routines into the model, not identifying them, so we look to alternative methods.

Giving device specific feedback to the user, especially if the family or carers have specific concerns is very important. Therefore, we view the data as multiple binary time series. We acknowledge there are many ways of viewing the data, however, this best fits Howz's goal and provides more specific information to relay to customers. Typical generalised linear models for binary data often use the Bernoulli distribution with the logit link function. Examples in the literature include; a Bernoulli autoregressive moving average model (Angelo Milani et al., 2019) and analysing eye tracking data with multilevel logistic regression (Barr, 2008). However, to accurately estimate the probability, the method must be able to follow the steep spikes and dips seen in the data (see Figure 3.1.2), which suggests classic linear modelling will not suffice. As indicated above, the current literature, contains time series models with autoregressive components. An alternative approach is to use Hawkes Processes. Hawkes processes place emphasis on using the recent data to inform the current probability on data that is more prone to steep spikes. Engelhard et al. (2018) use Hawkes processes for predicting smoking events and Jewell et al. (2019) for identifying spikes in calcium imaging data.

We propose a new model specification for modelling a sensor that takes inspiration from Hawkes processes within a generalised linear framework. It will dynamically vary the probabilities that a sensor will be triggered throughout the day, based on the prior data from the sensor itself and the sensors around it to incorporate the effect of routines in the data. In Section 3.2 we present the method and in Section 3.3 we apply it to Howz data. In Section 3.4 we compare this method to a logistic regression model and in Section 3.5 we summarise the results and discuss potential avenues for further research.

3.2 Model

The following method considers each sensor separately and models the behaviour as a binary time series conditional on the data up until the previous time point. Recall the data is sampled every 15 minutes, where T is number of data samples we have. Similar to the set up in Fokianos and Moysiadis (2017), and Milani et al. (2018), let $\{Y_t \ : \ t = 1, \ldots, T\}$ be a binary time series, the population random variable we are seeking inference upon. Then let y_t be the observed binary outcome for the sensor of interest at time t:

$$y_t = \begin{cases} 1 & \text{If the sensor of interest is triggered within the 15 minutes immediately} \\ & \text{prior to time } t, \\ 0 & \text{Otherwise.} \end{cases}$$

Note that, if the sensor is triggered multiple times in the 15 minutes this information is not included.

Let F_{t-1} be the information we have until time t,

$$F_{t-1} = \{y_{t-1}, \ldots, y_1, z_{j,t-1}, \ldots, z_{j,1}, p_{t-1}, \ldots, p_1\}$$

where y_i for $i < t$ from the modelled sensor. We let J denote the number of other sensors and $z_{j,i}$ be the other sensors $j \in J$ at time $i < t$ and prior expectations $p_t = \mathbf{E}(Y_t|F_{t-1})$. The conditional distribution for each Y_t is modelled as,

$$Y_t|F_{t-1} \sim \text{Bernoulli}(p_t).$$

Using a generalised linear model framework, we estimate the conditional mean, p_t, which is the probability that $Y_t = 1$ at time t given the information up until time t. In order to assess whether there has been a change in behaviour, the previous information for a household's routine is very important.

Motivated by our application we now consider the form of the linear predictor, Δ_t. For the Bernoulli distribution, a natural choice for a canonical link function is the logit function.

$$\Delta_t = \text{logit}(p_t) = \log\left(\frac{p_t}{1-p_t}\right).$$

This maps the linear predictor using the link function from $(-\infty, \infty)$ to $(0, 1)$ for the estimated probability p_t. Due to the placement of sensors throughout the house, they will often be triggered in a certain order, i.e. to get from the bedroom to the bathroom, a member of the household's must go through the hallway. We would also expect certain sensors to have some dependence on when it was last seen. For example, the hallway could be triggered before entering the bathroom and exiting, or if the kettle is used we may not expect to see the kettle again for a few hours.

We also include a seasonal factor due to sensors often being seen at similar times days previously. For example; putting on the kettle and having a morning coffee before leaving for work. Therefore, Δ_t should contain an autoregressive term b_t, a dependence on other sensors $c_{j,t}$ and a seasonal factor d_t:

$$\Delta_t = a + b_t + \sum_{j=1}^{J} c_{j,t} + d_t.$$

Moysiadis and Fokianos (2014) let $b_t = \eta b_{t-1} + \nu y_{t-1}$ where $\eta, \nu \in \mathcal{R}$. However, due to the steep changes in probability, see Figure 3.1.2, we need to place more restrictions on η. Instead we take take inspiration from Jewell et al. (2019) who include a rate parameter, η, that decays exponentially. In the context of the sensor dataset, we would want a spike to occur for $b_t, c_{j,t}$ and d_t based on the previous time point. For this application this could be in a negative or positive direction depending on the sensor, i.e. if a kettle sensor is triggered we may not expect the next one for a few hours, however a lounge sensor might be triggered a lot if the household spends most of the day there. Take b_t such that

$$b_t = \phi_b b_{t-1} + \pi_b y_{t-1},$$

where $\pi_b \in \mathcal{R}$ is the spike in the estimate. The equation will decay exponentially at a rate governed by ϕ_b where $|\phi_b| < 1$ to incorporate positive and negative rates of decay.

We express the other sensors similarly,

$$c_{j,t} = \psi_j c_{j,t-1} + \tau_j z_{j,t-1} \quad \text{for sensors } j = 1, \ldots, J.$$

For the seasonal term, d_t, we would like to acknowledge that routines can vary in their timing. Thus we have allowed 15 minutes either side of the time the previous day (45 minute window),

$$d_t = \phi_d d_{t-96} + \pi_d \max(y_{t-96-1}, y_{t-96}, y_{t-96+1}).$$

The term y_{t-96} signifies the same time the day previously, due to having 96 points a day as the data are sampled every 15 minutes. With this formulation of Δ_t we have,

$$p_t = \frac{\exp(a + b_t + \sum_{j=1}^{J} c_{j,t} + d_t)}{1 + \exp(a + b_t + \sum_{j=1}^{J} c_{j,t} + d_t)}.$$

Now we have defined p_t, we can simplify for ease of parameter estimation. We are able to simplify the link function in the following proposition.

Proposition 3.2.1. *We can simplify b_t to;*

$$b_t = \pi_b \sum_{i=1}^{t-1} \phi_b^{i-1} y_{t-i}.$$

Similarly,

$$c_{j,t} = \tau_j \sum_{i=1}^{t-1} \psi_j^{i-1} z_{j,t-i}.$$

For the seasonal term, considering $t \geq 96$,

$$d_t = \pi_d \sum_{i=1}^{\lfloor \frac{t}{96} \rfloor} \phi_d^{i-1} max(y_{t-96i-1:t-96i+1}) \qquad for\, t \geq 96.$$

Therefore, Δ_t and p_t can be expressed as,

$$\Delta_t = a + \pi_b \sum_{i=1}^{t-1} \phi_b^{i-1} y_{t-i} + \sum_{j=1}^{J} \tau_j \sum_{i=1}^{t-1} \psi_j^{i-1} z_{j,t-i}$$
$$+ \begin{cases} \pi_d \sum_{i=1}^{\lfloor \frac{t}{96} \rfloor} \phi_d^{i-1} max(y_{t-96i-1:t-96i+1}) & if\, t \geq 96, \\ 0 & Otherwise. \end{cases},$$

$$p_t = \frac{exp(\Delta_t)}{1 + exp(\Delta_t)}.$$

For a detailed proof see Appendix B.1.

We estimate the parameters using the Maximum Likelihood.

Proposition 3.2.2. *Let*

$$\alpha = \{a, \pi_b, \phi_b, \tau_1, \psi_1, ..., \tau_j, \psi_j, \pi_d, \phi_d\},$$

be the parameters we wish to estimate. Then we can formulate the log-likelihood as,

$$l(\alpha|\{y_t\}_{t=1}^{T}) = \sum_{t=1}^{T} y_t \Delta_t \; - \; \sum_{t=1}^{T} log\,(1 + exp(\Delta_t))$$

For the full derivation see Appendix B.2.

Due to the form of Δ_t, analytical expressions for $\hat{\alpha}$ are not attainable. Thus to estimate $\hat{\alpha}$, we use a common numerical optimization method called Broyden–Fletcher–Goldfarb–Shanno (BFGS), see Fletcher (1987). This optimisation method and versions of it are widely used in many research areas to find a local minimum of a differentiable function (Ge et al., 2018; Zhang et al., 2020; Wan et al., 2020). We provide the derivatives for the log-likelihood, $\nabla l(\alpha)$, which are given in Appendix B.3 for reference. Due to the logarithms within the gradient function, the iterative optimization algorithms are prone to exploding gradients. This often occurs in neural networks (Pascanu et al., 2013). One way to reduce the occurrence of this problem is by using an adaptive step size. We use 6 different step sizes at each iteration, taking the best choice at each step to reduce the occurrence of this problem for the results in Section 3.3.

Now we know how to fit the model, we need to perform model selection to decide which and how many terms to include within Δ_t. In the model set up, we have

2 parameters for each sensor plus the constant term. However, some sensors (even the regression on the sensor of interest) may not be important for modelling the probability. In the interest of parsimony we use a stepwise regression selection process to select which sensors to include.

We use a greedy forward selection process, adding a sensor to the model if it improves the model for a certain criterion, a common method of fitting regression models (Draper and Smith, 1998). Bayes Information Criterion (BIC) and Akaike's Information Criterion (AIC) are well-used model selection criteria to test the goodness of fit for each regression model in the selection process. Letting $k = |\alpha|$ be the number of parameters, BIC is defined as $k\log(T) - 2l(\alpha)$. AIC is similarly defined, however it replaces $k\log(T)$ with $2k$. We choose to use BIC as it is a more conservative criteria than AIC, penalising more for adding extra parameters. This is preferable for our application as we want to reduce the computation time as we are fitting the model to each sensor and across customers. We now provide results on different sensors provided by Howz.

3.3 Howz Data Example

Due to the categorical nature of the data, it is difficult to identify the accuracy of each estimated probability in comparison to 0-1 sensor triggering. Especially when there is often a low chance of a sensor triggering. Therefore, to assess the validity of the model, we construct point-wise quantile intervals over a month of data. The data provided by Howz comprises of two months, January and February. We take data from January to estimate the parameters, $\hat{\alpha}$, using the method described in Section 3.2. Then, we run the model on February, estimating the one-step ahead probabilities, using the model parameter estimates from January.

We choose 6 different step sizes, $\{5 \times 10^{-5}, 0.01, 0.5, 1, 2.5, 4\}$ to give a wide range for the adaptive BFGS. For the stopping criterion we set $|l(\alpha_n)| < 1e - 5$ for the nth step in the BFGS and set a maximum number of iterations to be performed as 500.

We take the data from February and simulate an online process, calculating the one-step ahead estimated probability, using the data gathered up until each future time t.

$$P(Y_{t+1}|F_t) = E(Y_{t+1}|F_t) = \frac{\exp(\hat{\Delta}_{t+1})}{1 + \exp(\hat{\Delta}_{t+1})} \tag{3.3.1}$$

$$\text{where} \quad \hat{\Delta}_{t+1} = \hat{a} + \hat{\pi}_b \sum_{i=1}^{t} \hat{\phi}_b^{i-1} y_{t-i+1} + \sum_{j=1}^{J} \hat{\tau}_j \sum_{i=1}^{t} \hat{\psi}_j^{i-1} z_{j,t-i+1} \tag{3.3.2}$$

$$+ \begin{cases} \hat{\pi}_d \sum\limits_{i=1}^{\left\lfloor \frac{t+1}{96} \right\rfloor} \hat{\phi}_d^{i-1} \max(y_{t-96i:t-96i+2}) & \text{if } t+1 \geq 96, \\ \\ 0 & \text{Otherwise.} \end{cases} \tag{3.3.3}$$

The first 96 estimates, the first day, are discarded as burn in. Now for each 15 minute period in 24 hours, we have around 27 estimated probabilities. While we are aware of the routines present in the data, it would be unrealistic to assume these groups of probabilities were identically distributed, i.e. a member of the household follows the same routine at the exact same time each day. Therefore, we use the Poisson Binomial Distribution to get the quantiles for our estimates. The Poisson Binomial distribution is the sum of independent Bernoulli trials which are not enforced to be identically distributed (Wang, 1993). This provides an interval closer to reality as we cannot assume the different probabilities are identically distributed but they are likely to be similar across days.

Let the estimated probabilities be $\{\hat{p}_{i,w}\}$ for $i = 1, \ldots, 27$ and w be the time period in the day, i.e. 1 , . . . , 96. Then the distribution can be expressed as,

$$\mathbf{X}_w = \sum_{i=1}^{n} X_{i,w} \quad \text{with} \quad X_{i,w} \sim \text{Bern}(p_{i,w}).$$

The mean and variance are,

$$\mu = \sum_{i=1}^{n} p_{i,w} \quad \text{and} \quad \sigma^2 = \sum_{i=1}^{n} p_{i,w}(1 - p_{i,w}).$$

The package `poibin` in R (Hong, 2013), provide estimates of the quantiles of this distribution (2.5% and 97.5%). To compare to the observed data, we take the binary time series and for each time period, 1 , . . . , 96 , sum the occurrences of the sensor of interest over February. If the model is appropriate we would expect the empirical sum to lie within the predicted quantile interval from the model.

We show one example from 2 different households; one with 10 sensors and one with 4 sensors. For more examples see Appendix B.4. The following two households are chosen from a range of households provided by Howz as they show different features. We seek to demonstrate that the method works where households have a small or large number of additional sensors and selected households that show sensors that have some "spikes" rather than a flat average behaviour over the day (like household 3 in the appendix). These are typical examples of households within the wider dataset provided to us.

3.3.1 Household 1

We present a household monitored by Howz, including 10 sensors; Bedroom, Lounge, Bathroom, Hallway, Kitchen, Fridge Door, Kettle, Front Door, Back Door and Toaster. We model the Bedroom sensor, using the model selection process in Section 3.2. The model selection indicates that the Hallway and Front Door sensors are informative for modelling the Bedroom sensor, along with the Bedroom sensor itself and the daily seasonal component. The parameter estimates are given in Table 3.3.1.

α	π_b	ϕ_b	$\tau_{Hallway}$	$\psi_{Hallway}$	$\tau_{FrontDoor}$
-3.009	1.578	-0.040	0.683	0.547	-0.403
$\psi_{FrontDoor}$	π_d	ϕ_d			
0.926	0.322	0.824			

Table 3.3.1: Parameter estimates from fitting the model in Section 3.2 to Household 1, Bedroom sensor.

Figure 3.3.1 depicts the 95% quantile band along side the number of true events over the 27 days which shows that the method adapts well to the change in the number of events over the day. The quantile band follows the clear peaks at the beginning of the household's day and the end, between 9am and 12pm and after 8pm, respectively. It is also able to follow the lower level activity throughout the afternoon. We have 96 points and 1 falls outside the quantile interval indicating that the false positive rate is controlled.

Table 3.3.1 shows that the Bedroom sensor has a large positive spike coefficient,

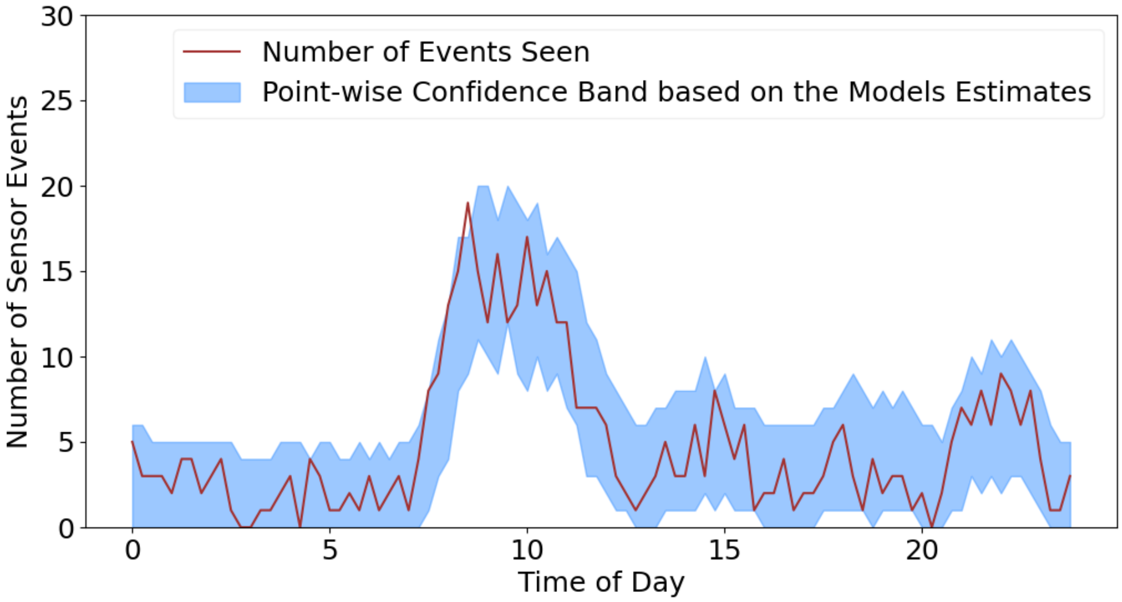


Figure 3.3.1: Number of events seen over 27 days (line) with the predicted 95% quantile interval (shaded band) for the Bedroom sensor, Household 1.

meaning that if we see the Bedroom sensor in the previous time period we are more likely to see it in the next time period. Conversely, if we see the Front door sensor, we are less likely to see the Bedroom sensor triggering in the next 15 minutes.

The computational time for this method is split into 3 parts; model selection, parameter estimation and probability estimates. In practice, the model selection and parameter estimation will be run offline. The model selection for the Bedroom sensor in this example took 424 seconds (machine with 4 cores, 2.8Ghz processor and 16Gb RAM) without parallelisation. The estimation of the parameters after model selection took around 21 seconds with 40 iterations of the adaptive BFGS (where each iteration checked 6 step sizes). Finally it took around 2 seconds to estimate the probabilities for all of February. As there are 96 probabilities estimated each day and 28 days of data, this shows the method is computationally efficient at calculating the probabilities.

3.3.2 Household 2

Consider a different household with 4 sensors; Bedroom, Hallway, Kettle and Microwave. To test how well this method works with varying amounts of data, this household has fewer sensors so will have vastly fewer covariates (the sensors being triggered). We wish to estimate the quantile interval over a day for the Bedroom sensor in February. The model selection returns that Kettle is informative with the autoregressive term and the daily seasonal term. The method estimates the parameters as in Table 3.3.2.

α	π_b	ϕ_b	τ_{Kettle}	ψ_{Kettle}	π_d	ϕ_d
-3.325	0.770	0.170	-2.124	0.310	0.637	0.817

Table 3.3.2: Parameter estimates from fitting the model in Section 3.2 to Household 2, Bedroom sensor.

Using these parameters we estimate the quantile interval displayed in Figure 3.3.2 and see again it follows the peaks seen in the total number of events. We can see this household tends to get up between 6am and 9am, with the user more likely to be active in the bedroom closer to 9am. The quantile band is able to follow this noisy upwards trend in the morning and steep decline around 9am, as well as the large spike around 10pm for the end of the household's day. Again the method is shown to control the false positive rate well with 3 outside the band.

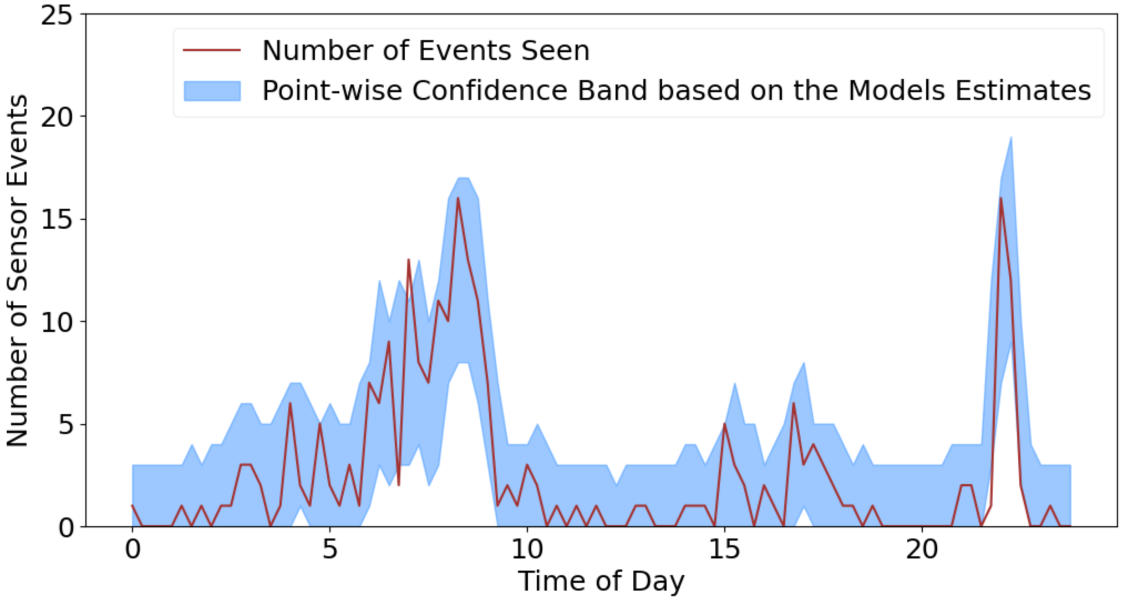


Figure 3.3.2: Number of events seen over 27 days (red) with the predicted 95% quantile interval (blue) for the Bedroom sensor, Household 2.

It is interesting to note that the constant term, a, is negative for all of the examples. This is due to balancing the rate and spike parameters, ensuring that the baseline probability of a sensor being triggered is small if none of the sensors chosen from the model selection have triggered. We also see from Table 3.3.2 that the Kettle sensor has a negative spike parameter. This tells us, if we see the kettle in the previous 15 minutes we are unlikely to see the Bedroom sensor. However for the Fridge Door and seasonal component, if we see it in the previous 15 minutes we are likely to see it in the next 15 minutes.

Household 2 has a smaller number of sensors, hence the computational time for this example is reduced. It took around 72 seconds for the model selection, 6 seconds for the parameter estimation with 25 iterations and 1 seconds to calculate the estimated probabilities for February.

3.4 Comparison to Logistic Regression

Next we compare to the logistic regression to further test the validity of the model. We define the linear predictor, Δ_t, using the same notation,

$$\Delta_t = a + \pi_b y_{t-1} + \sum_{j=1}^{J} \tau_j z_{j,t-1} + \pi_d \max(y_{t-96-1:t-96+1}).$$

The package `statsmodel` (Seabold and Perktold, 2010) in python provides the estimated one-step ahead estimated probabilities. These probabilities can be used to form similar quantile intervals as the examples in Section 3.3.

3.4.1 Howz Data Examples

Using these probabilities we present how many points fall outside the quantile interval for the Howz data examples in Section 3.3 and Appendix D. Table 3.4.1 shows the Bernoulli autoregressive method has fewer points outside the 95% quantile interval for all households. See Appendix B.4 for the other household examples.

For a more direct visual comparison we present the quantile interval figure for the Bedroom sensor from Household 2 using the logistic model, in Figure 3.4.1 . Here it is clear to see the benefits of the Bernoulli Autoregressive Model, Figure 3.3.2, over the Logistic Model. When there are sharp spikes and dips (in the morning and evening), the Logistic Regression Model struggles to follow the behaviour. It also struggles with fast changes in behaviour, e.g. between 5 and 10am, the logistic method slowly trends upwards. In comparison, in Figure 3.3.2 we can see the method is following

		Logistic	Bernoulli Autoregressive Model
Household	Sensor	Outside QI	Outside QI
1	Bedroom	9	1
1	Fridge Door	8	2
2	Bedroom	7	3
2	Kettle	4	0
2	Hallway	12	5
2	Microwave	6	0
3	Kettle	4	0

Table 3.4.1: Comparison between logistic model and the Bernoulli Autoregressive model with columns (left to right); household number, type of sensor being predicted, how many points lie outside of the quantile interval for the Logistic Regression Model and the Bernoulli Autoregressive Model.

the behaviour much better. Further, it can be seen that it is able to follow the sharp spikes in behaviour well.

3.4.2 Simulating from the Bernoulli Autoregressive Model

Alternatively we can simulate a new sensor X to compare our method to the Logistic Regression. To set up the simulation, we imagine a new sensor within Household 2. Sensor X is dependent on the Bedroom sensor, the sensor itself and the seasonal

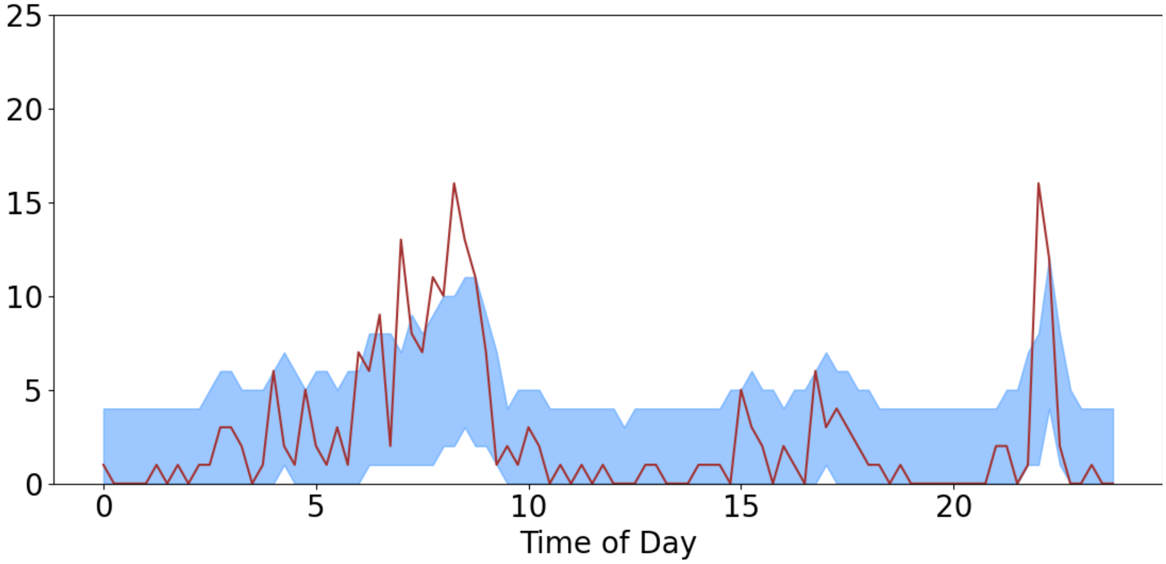


Figure 3.4.1: Number of events seen over 27 days (red) with the predicted 95% quantile interval (blue) for the Bedroom sensor, Household 2, using Logistic Regression Model.

component. To simulate from the Bernoulli Autoregressive model, we specify the parameters in Table 3.4.2.

α	π_b	ϕ_b	$\tau_{Bedroom}$	$\psi_{Bedroom}$	π_d	ϕ_d
-3.3	0.3	0.5	0.5	0.9	0.4	0.8

Table 3.4.2: Parameters to Simulate Sensor X from the Bernoulli Autoregressive Model.

In context, the positive spike parameters mean we are more likely (to varying degrees) to see sensor X triggered if we have seen the Bedroom and the sensor itself recently, as well as seeing the sensor at a similar time the day before. Using the

Bedroom sensor observations from household 2, we can simulate January and February for Sensor X using our method. We chose not to run the model selection so we can directly compare the two models. Hence we estimate the parameters chosen above on the January data for both models. Then similar to previously, run each method to get the probability estimates for February. We repeat this experiment 500 times to gather the mean number of points outside the quantile intervals.

Figures 3.4.2 and 3.4.3 show one realisation from each model, we can see that our method is better able to follow the large spike between 5am and 1pm. There are 6 points outside the quantile interval for our method versus 12 in the Logistic model. Overall from the simulations, the Bernoulli model has a mean of 2.26 outside the quantile interval whereas the Logistic model has a mean of 10.32. This shows that the Bernoulli model is better able to predict the simulated data in this example.

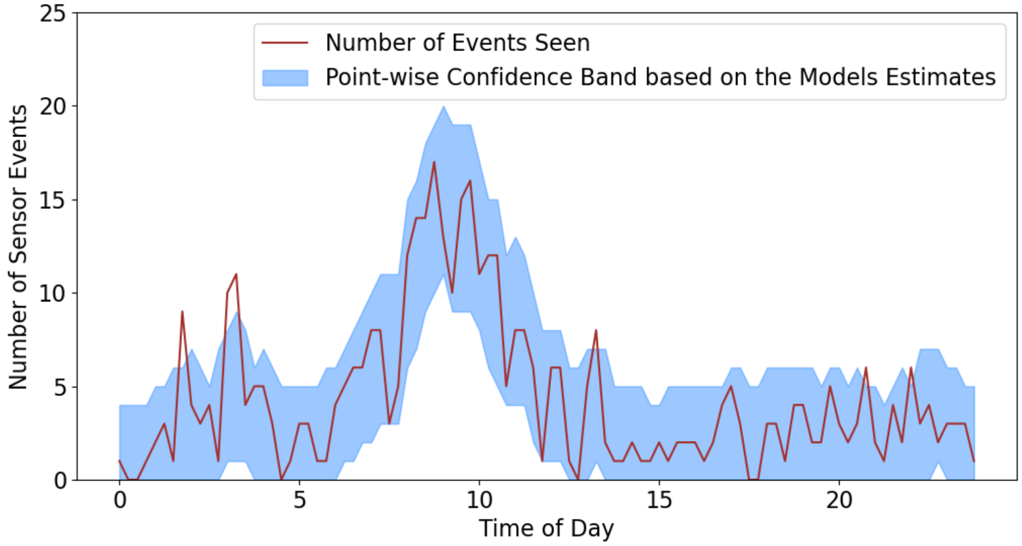


Figure 3.4.2: Number of events seen over 27 days (red) with the predicted 95% quantile interval (blue) for Sensor X, for a realisation using the Bernoulli Autoregressive Model.

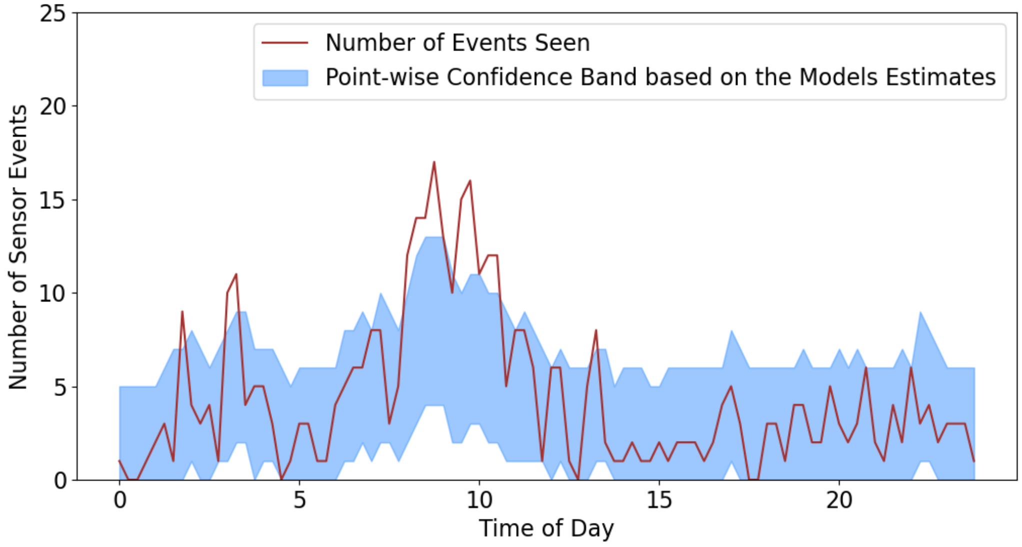


Figure 3.4.3: Number of events seen over 27 days (red) with the predicted 95% quantile interval (blue) for Sensor X, for a realisation using Logistic Regression Model.

3.4.3 Simulating from the Logistic Model

Using the same set up as Section 3.4, we instead simulate from the Logistic Regression model using alternative parameters in Table 3.4.3 for a new sensor Y. Again we repeat the experiment 500 times.

α	π_b	$\tau_{Bedroom}$	π_d
-2.8	0.8	1.2	1.4

Table 3.4.3: Parameters to Simulate Sensor Y from the Logistic Regression.

From the simulations, the Bernoulli model has a mean of 2.97 outside the quantile interval whereas the Logistic model has a mean of 2.99. This confirms our method is able to predict the behaviour as well as the Logistic regression. For an example of what Sensor Y could look like, Figures 3.4.4 and 3.4.5 show that both methods are able to follow the spike just before 10 and both follow the true number events seen well throughout the day. There are 3 points outside the quantile interval for our method versus 5 in the Logistic regression. Overall it is clear to see that our method performs well for our application, when simulating from our model and when using simulated data from the Logistic Regression model.

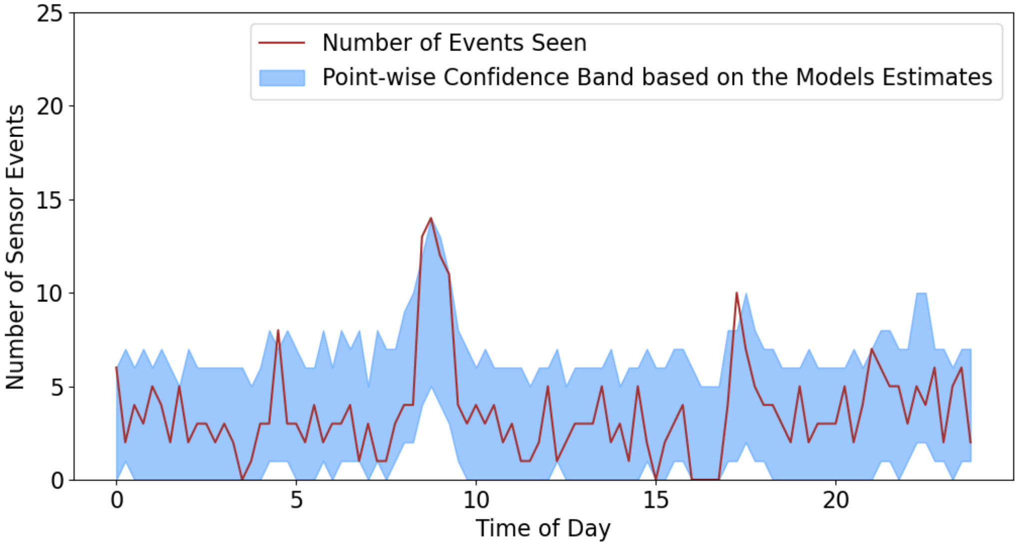


Figure 3.4.4: Number of events seen over 27 days (red) with the predicted 95% quantile interval (blue) for Sensor Y, for a realisation using the Bernoulli Autoregressive Model.

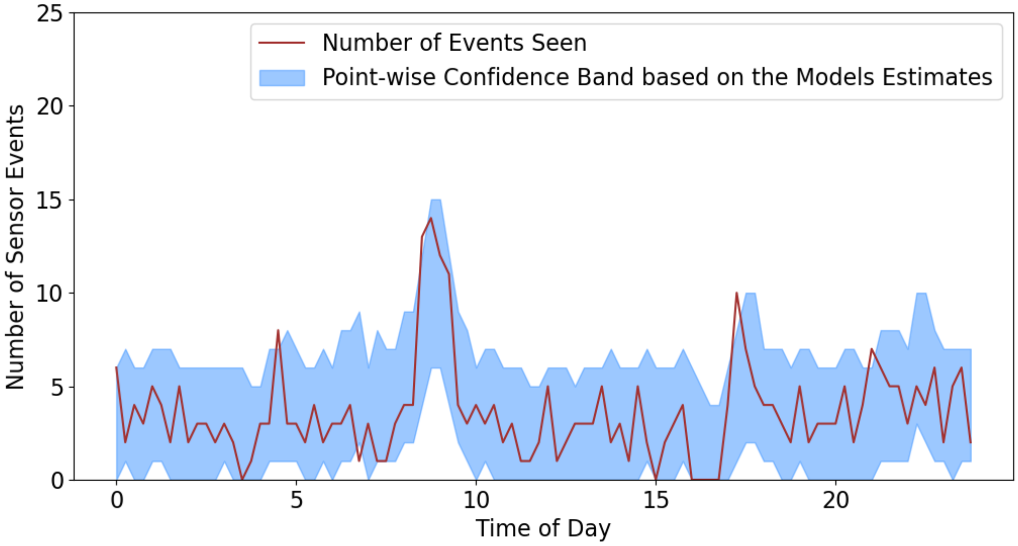


Figure 3.4.5: Number of events seen over 27 days (red) with the predicted 95% quantile interval (blue) for Sensor Y, for a realisation using Logistic Regression Model.

3.5 Conclusion and Future Research

In this paper we have presented a method which is able to predict the one step ahead probability of a sensor being triggered given other household sensors. Using model selection, it is able to choose which sensors are of importance to the estimation and provide a good estimate on the Howz data as seen in Section 3.3. After the initial model selection and parameter estimation, it is computationally efficient enough to be run online or at the end of every day.

We have shown through the use of quantile intervals the validity of the method with the number seen outside the quantile band are controlled well, with the results able to follow the spikes and dips throughout the 24 hour period. When comparing to the Logistic Regression, we see that the method is able to follow the spikes within the data better with fewer outside the quantile intervals. The model adapts well to households with a range of different sensors, as well as households with between 2 and 10 sensors. When simulating data from our model, we can see the method outperforms the logistic regression. It is also able to perform as well as the Logistic regression when simulating data from a Logistic Regression model.

The next step would be to adapt this to identify when changes in behaviour occur on a 15 minute/daily basis. Currently the data presented is assumed to be regular for the household, i.e. no significant changes to household's routines and behaviours. However, this method can currently provide information and alerts on a monthly scale, when a household is outside the quantile intervals, indicating a potential change. Currently this method focuses on predicting each sensor separately, this method could also be extended to look at the multivariate fit of the sensors to further consider the interactions between the sensors.

Chapter 4

Framework to Identify Subtle Changes in Probabilistic Cluster Membership: an application to passive household monitoring

4.1 Introduction

By 2039, it is projected that one in four people in the UK will be aged 65 or over (ONS, 2021), in part, this is believed to be due to increased life expectancy and other factors. While this is a cause for celebration, the pressure on the NHS and other health care providers to help an ageing population will increase. As of 2019, close to a quarter of those above the age of 65 are found to be living by themselves (Age UK, 2019). This can result in, amongst other effects, more trips to the emergency room

(Dreyer et al., 2018). To find new ways to help improve health and care of the older adults, people are increasingly turning towards technology.

One such innovation is Howz, a product which places non-intrusive sensors within the homes of elderly people and care facilities. The data gathered from these sensors is used to observe changes in a household's behaviour. A possible example is if a customer loses confidence in using kitchen appliances over a period of time, transitioning from cooking food using the oven to microwave meals. Identifying these changes can be used to improve the health and well-being of the households. Once Howz has detected a change, they inform the household, family member or carer, to prompt awareness and conversations around well-being.

Howz is in a period of rapid growth with a large number of households adopting the Howz sensors. Deciding whether households follow similar patterns to others can be a valuable tool for prediction. Many changes in human behaviour are subtle and so detecting deviations from households they were once similar to can be a way of identifying change. An example is the dynamics within care homes, if a proportion of residents deviate from their routines in a similar way this can help inform the staff around the way they are caring for the residents. In this paper we investigate the group structure across multiple households and use this to identify changes within an individual household's behaviour over time.

4.1.1 Howz Anonymised Dataset

The data collected by Howz are from static infra-red motion sensors, door sensors and smart plug sensors placed within the home. For example; Bedroom door, Hallway

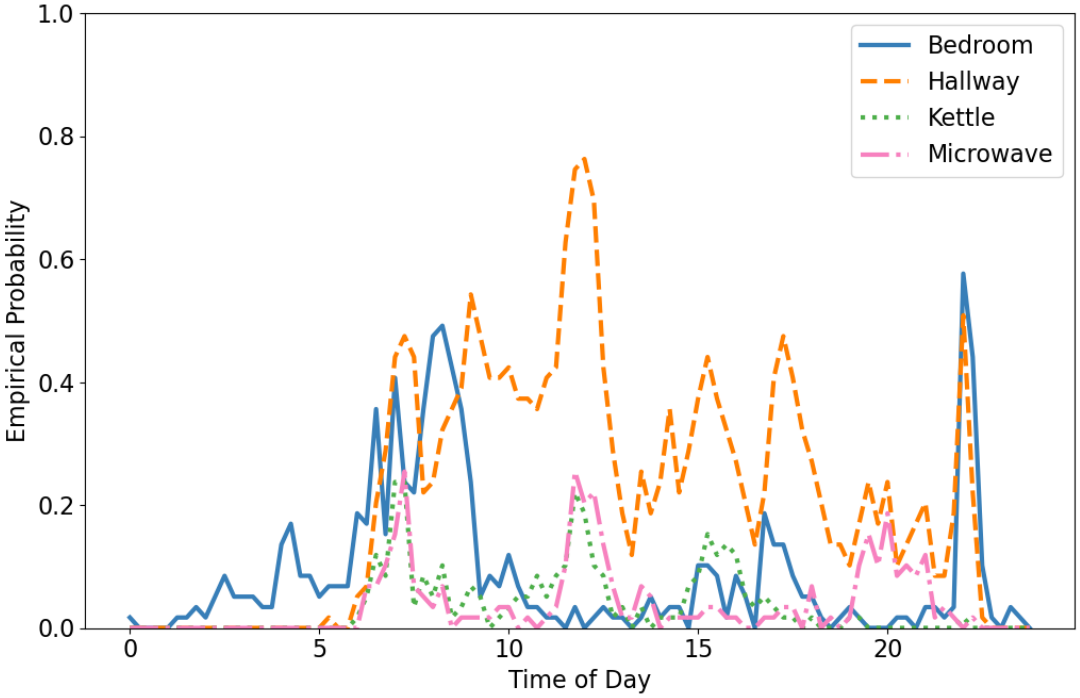


Figure 4.1.1: Empirical probabilities for a household's sensors being active in a 15 minute interval across the day as explained in Section 4.1.1. The sensors shown are Bedroom, Hallway, Kettle and Microwave.

motion and Kettle plug sensors. The data gained from each labelled sensor is a binary observation with the timestamp of activation. Howz typically provide 3 sensors as a starter package for a new customer, with the option to install additional sensors at a later point. It is up to the customer how many sensors are used with some households having up to 10 sensors. Due to information only being received through interaction with the sensors, Howz often aggregate data into 15 minute periods across the day. In Figure 4.1.1 we present the probability that a sensor was triggered in a 15 minute time interval. We estimate the probabilities empirically using 6 weeks of data.

4.1.2 Statistical Motivation and Structure

For this application, we aim to monitor households who behave similarly across time, to identify when changes occur. A popular method to group similar objects is clustering. There exists a number of different clustering algorithms in literature with frequent comparative reviews (Jain et al., 1999; Kinnunen et al., 2011; Rodriguez et al., 2019). These methods are often split into two groups; hard and soft clustering. Common hard clustering methods include k-means and DBSCAN. These methods take an item and output an estimated cluster label for the cluster it is most similar to. Using these methods, if we want to identify changes in cluster membership over time we would monitor the household cluster label. However, if cluster membership is uncertain, random noise would result in many changes in cluster membership.

Instead we focus on soft clustering methods, such as Gaussian Mixture Models, able to output the probability of being in each cluster at each time step. To motivate this, consider Figure 4.1.2 which shows varying behaviours across different household's Kettle sensors. Let us assume pre-existing knowledge of the following clusters; morning activity, morning and afternoon activity, and low level activity. Due to household variation we might see day to day, Household 4 and 5 could move between the first two clusters frequently. However, these would not reflect true changes in behaviour as they are simulated from the same underlying distribution. Hence, by using soft clustering methods, if households are consistently similar to two clusters, we will be less likely to identify changes which are due to random variation.

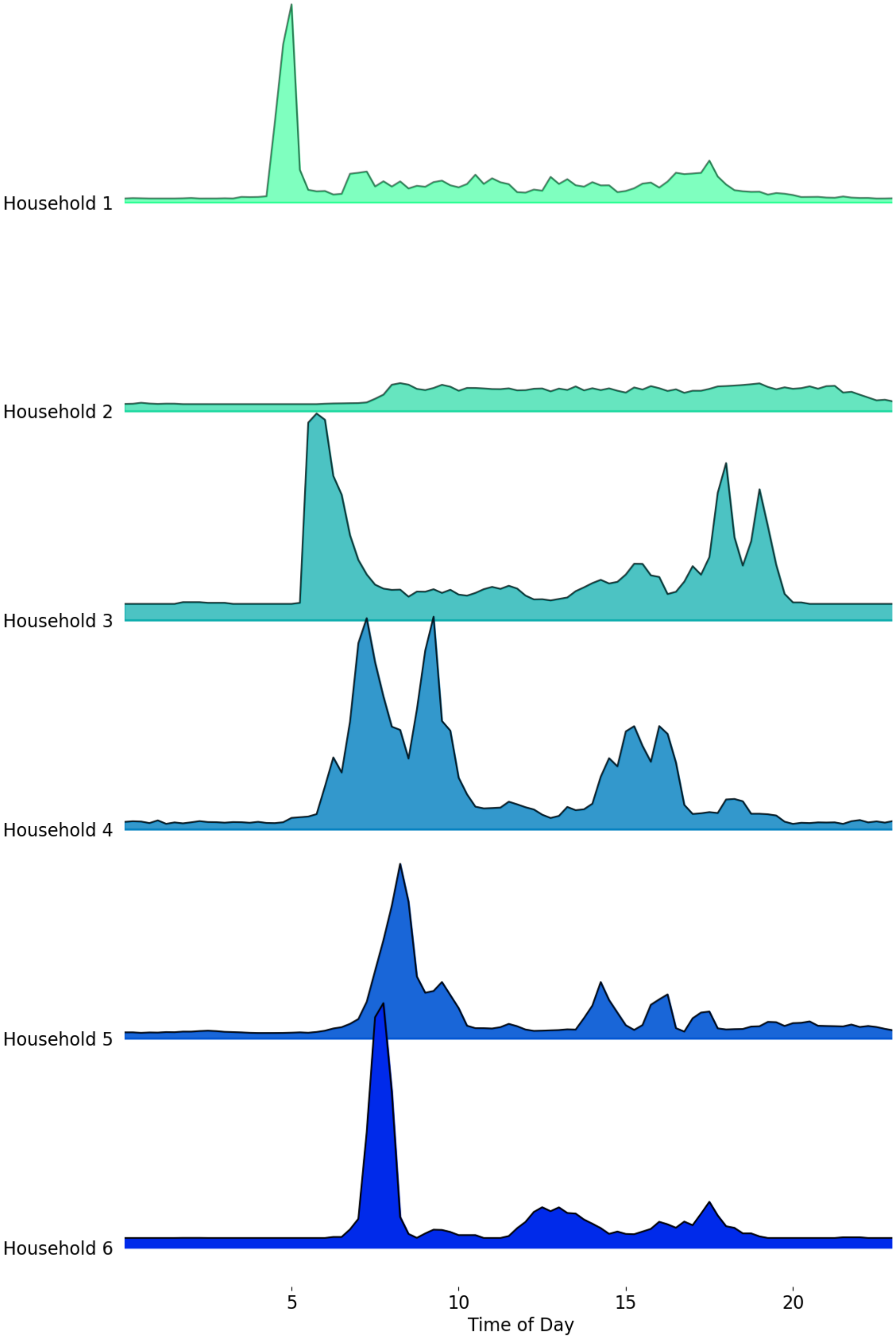


Figure 4.1.2: Average empirical probabilities for a Kettle sensor triggering in 6 households over 6 weeks.

Clustering methods alone can not tell us the significance of changes in a household's cluster membership probabilities throughout time. One common way to formally detect changes in behaviour within time series data is changepoint detection. Changepoint detection aims to find the point in time at which the data generating process changes to an alternative one. This is a well-studied area (Killick et al., 2012) with an overview found in Eckley et al. (2011). The standard algorithms to detect changes in mean for Normal distributed data are not suitable for soft cluster membership probabilities, due to the categorical nature and the fact that the data are constrained between 0 and 1, and sum to 1. Therefore, we use a Dirichlet distribution assumption when adapting changepoint methods to handle this data.

Here we propose a new generalised framework to identify subtle changes in behaviour over time, in an offline approach. We combine clustering and changepoint techniques to provide the user with a framework to monitor when members of each cluster move across time and use changepoint analysis to identify if this change in cluster membership is a significant behaviour change. In Section 4.2 we detail the methods we have chosen for each step in our application followed by the generalised framework. In Section 4.2 we apply the framework to simulated data, and then to the Howz dataset in Section 4.3. In Section 4.4, we summarise the results and discuss future research.

4.2 Framework Set Up

In this section, we provide a combination of methods we have chosen for the Howz dataset with a simple example for how the framework could be used. In Section 4.2.1 we explain a pre-processing step unique to the Howz application. In Section 4.2.2 we introduce the algorithm used to extract the cluster membership probabilities. We then present the Dirichlet distributed changepoint algorithm used to identify when significant changes occur in Section 4.2.3. Finally, in Section 4.2.4 we summarise these steps and propose the generalised framework.

4.2.1 Initial Modelling for Howz

For our application, one way of identifying changes in behaviour within a group structure is to compare similar sensors across households, e.g. the Kettle sensor in Figure 4.1.2. In considering sensors separately, Howz are able to provide device specific feedback to the users. This analysis could then be replicated for different sensors of interest for the user.

To identify changes in behaviour over time, we consider the data as time series of sensor observations. One sensible approach to characterizing behaviour within a household, is to look at the probability of a sensor activating at a certain time. However, taking the empirical probability, as in Figure 4.1.1, will result in rolling averages of weeks or months of data to create a single data point. This is not ideal for our application, as any changes will be smoothed over time and we wish to identify changes on shorter time scales.

To combat this, we consider the data on a daily scale, i.e. each day is a data point in the time series. The complication here is that the raw binary data for a sensor are 0/1 observations, thus it is not clear what the probability of a sensor activation at a particular time is as we have a single observation (activated or not). Gillam et al. (2022) propose a novel binary time series forecasting approach and apply it to the Howz dataset. They provide a method to forecast the one-step ahead probabilities a sensor will be triggered in the next 15 minutes. A model is fit for each sensor and household independently. Hence similar to Figure 4.1.1, we can gain the estimated probability a sensor will trigger at 15 minute intervals, but on a daily scale instead of over 6 weeks. We choose to take these estimated probabilities as the data we will use within the framework for our application.

In Figure 4.2.1, we show the average predicted probability for two different household's Kettle sensor, using the method described in Gillam et al. (2022). We took 8 weeks of data, 30^{th} April - 24^{th} June. We estimated the parameters on the first two weeks of data and then predicted the one-step ahead probability for the rest of the data. This shows one household with a clear morning and afternoon routine, i.e. turning on the Kettle when they get up and having an afternoon cup of tea. The other household has no clear routine, a smaller chance of triggering the Kettle throughout their day.

4.2.2 Calculating the Cluster Probabilities

We are proposing a method to identify changes in cluster probabilities, combining clustering and changepoint analysis. In doing so we assume that the clustering has

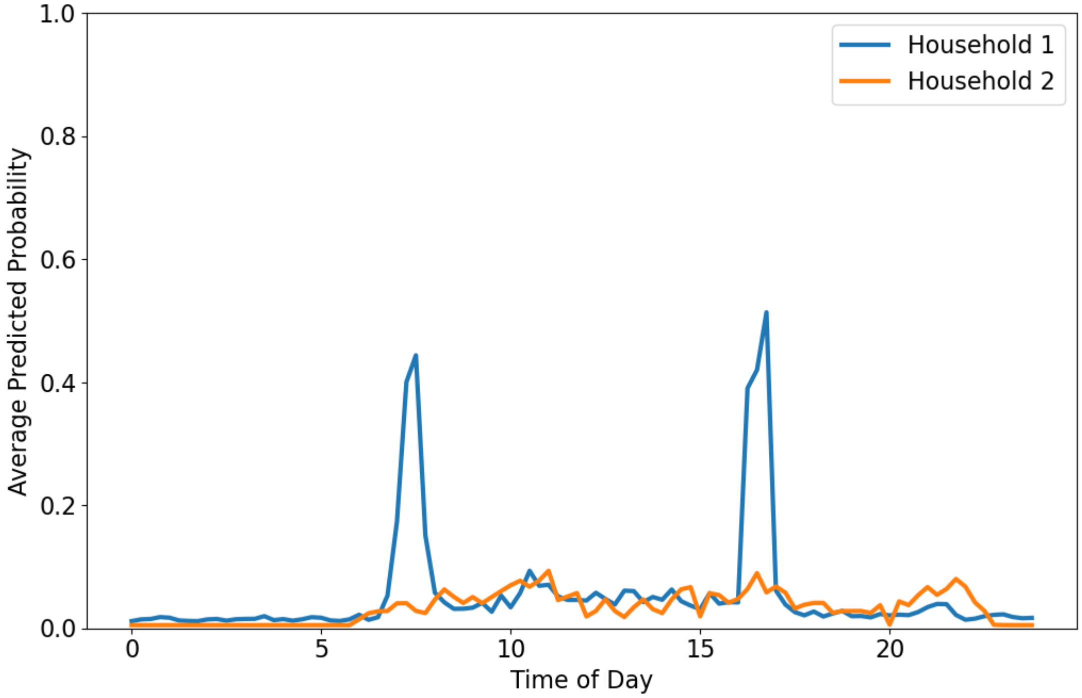


Figure 4.2.1: Two households with the average predicted probability over a week of the Kettle sensor triggering in the next 15 minutes.

taken place and we are now using those identified clusters to predict the membership of a new data point. Therefore, we do not discuss the estimation of the number of clusters and cluster attributes here. As a standalone problem, this is technically classification, but as we are utilising our known clustering method for the prediction we use the term clustering to make this clear. Importantly, for our application we can use expert knowledge from Howz to decide on the number of clusters and use information relevant to Howz goals to decide cluster attributes.

Due to the number of households and the differing behaviour of sensors (seen in Figure 4.1.2), we do not expect the clusters to have similar shapes. Given this we assume the clusters represent a mixture of Gaussian distributions as they allow for differing means and variances between clusters, as well as the ability to output the probability of belonging to each distribution.

Whilst we are assuming the clustering is known, Gaussian Mixture Models can be used to identify cluster attributes (McLachlan and Basford, 1988) and several methods for identifying the numbers of clusters are compared in Kodinariya and Makwana (2013).

We introduce the following notation for our clusters, let each day of data we wish to calculate the cluster probabilities of be $x_{i,t} \in \mathbb{R}^p$, for each household i and time $t \in 0, \ldots, T$. This results in a time series of observations, $x_i = \{x_{i,0}, \ldots, x_{i,T}\}$ for each household i. We can then define the density function for the mixture of K Gaussian distributions as

$$f(x_{i,t}) = \sum_{k=1}^{K} w_k \phi\left(x_{i,t}; \mu_k, \Sigma_k\right),$$

where

$$\phi\left(x_{i,t}; \mu_k, \Sigma_k\right) = \frac{1}{\sqrt{2\pi|\Sigma_k|}} exp\left(-\frac{1}{2}(x_{i,t} - \mu_k)^T \Sigma_k^{-1}(x_{i,t} - \mu_k)\right)$$

is a multivariate Gaussian distribution, with known mean $\mu_k \in \mathbb{R}^p$ and covariance matrix $\Sigma_k \in \mathbb{R}^{p \times p}$. The weight w_k is the probability of selecting cluster k where $\sum_{k=1}^{K} w_k = 1$. Let $z_{i,t}$ be the cluster label for $x_{i,t}$. Hence to find the probability of belonging to each cluster, we consider the conditional distribution of being in the cluster $z_{i,t}$ given the data $x_{i,t}$ given by,

$$\begin{aligned} p(z_{i,t} = k|x_{i,t}) &= \frac{p(z_{i,t} = k)p(x_{i,t}|z_i = k)}{p(x_{i,t})} \\ &= \frac{w_k \phi(x_{i,t}; \mu_k, \Sigma_k)}{\sum_{m=1}^{K} \pi_m \phi(x_{i,t}; \mu_m, \Sigma_m)}. \end{aligned}$$

We simplify to $y_{i,k,t} = p(z_{i,t} = k|x_{i,t})$ as this is our time series for use in Section 4.2.3.

For illustrative purposes, we introduce a simplified simulation to apply each step of the framework to. We take the households in Figure 4.2.1 as inspiration for our cluster means in Figure 4.2.2. For visualisation purposes, we consider two-hourly observations. To simulate our households, we simulate 25 households from each cluster by adding Gaussian noise to the probabilities in Figure 4.2.2, constraining the probabilities to be between 0 and 1. We then designate household 1 to 25 as similar to Cluster 1 and 26 to 50 as similar to Cluster 2. These are our new population means for each household. Figure 4.2.3 represents an example of the population for Kettle sensors, showing the variation across households, that we wish to cluster.

Taking the cluster population means from Figure 4.2.2 and setting the cluster variances as 0.05 to allow for some reasonable overlap between the clusters, we simulate 31 new days and output the cluster membership probabilities for each day for each household. We set the weight for each cluster to be equal, 0.5. As we have two clusters, a probability of 0.5 or more will place a day in cluster 1 and a probability of less than 0.5 will result in placement into cluster 2. Figure 4.2.4 shows the method correctly placing each household from ‘Day 1’ similar to Household 1 in the first cluster and those similar to Household 2 in the second cluster. We can also see the probabilities vary household to household as expected from the varied data seen

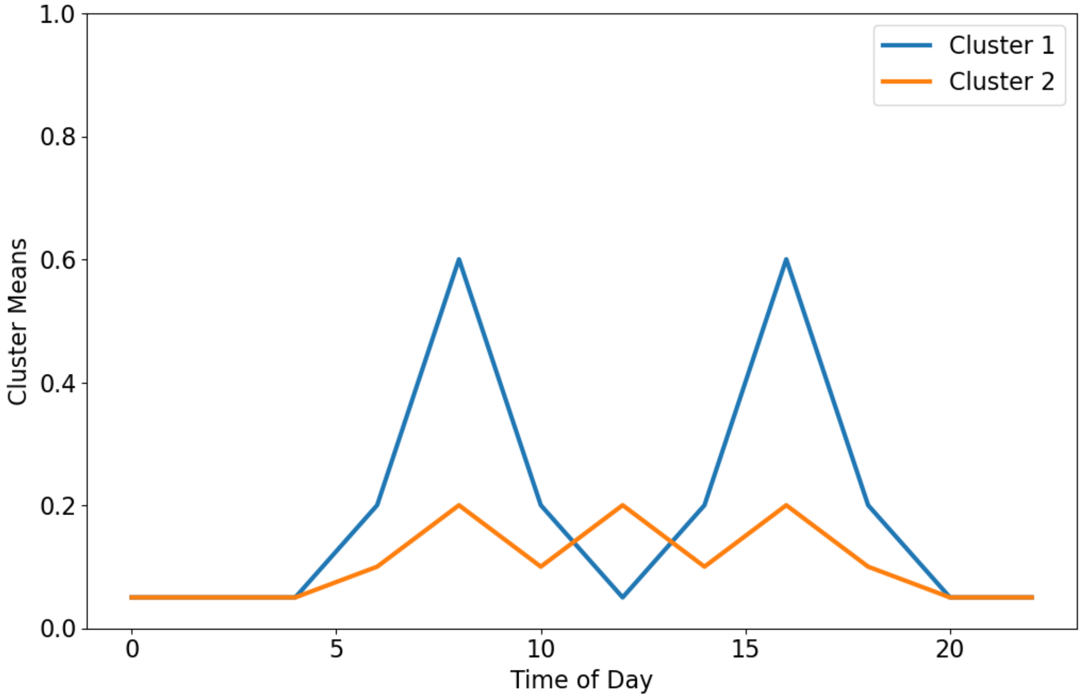


Figure 4.2.2: Two simple cluster means inspired by the data from Figure 4.2.1.

in Figure 4.2.3. An implicit assumption here is that the cluster means from Figure 4.2.2 remain the same across days as we do not consider the updating of these here.

We present 3 different scenarios simulated from the Kettle sensor probabilities in Figure 4.2.5 with the following behaviour over 31 days; Household A where the household has consistent behaviour not moving from cluster 1. Household B where the household stops using the Kettle so regularly in the morning and afternoon. Finally Household C who lies closer to the boundary between the two clusters, e.g. uses the Kettle in the morning and afternoon to a lesser degree than A.

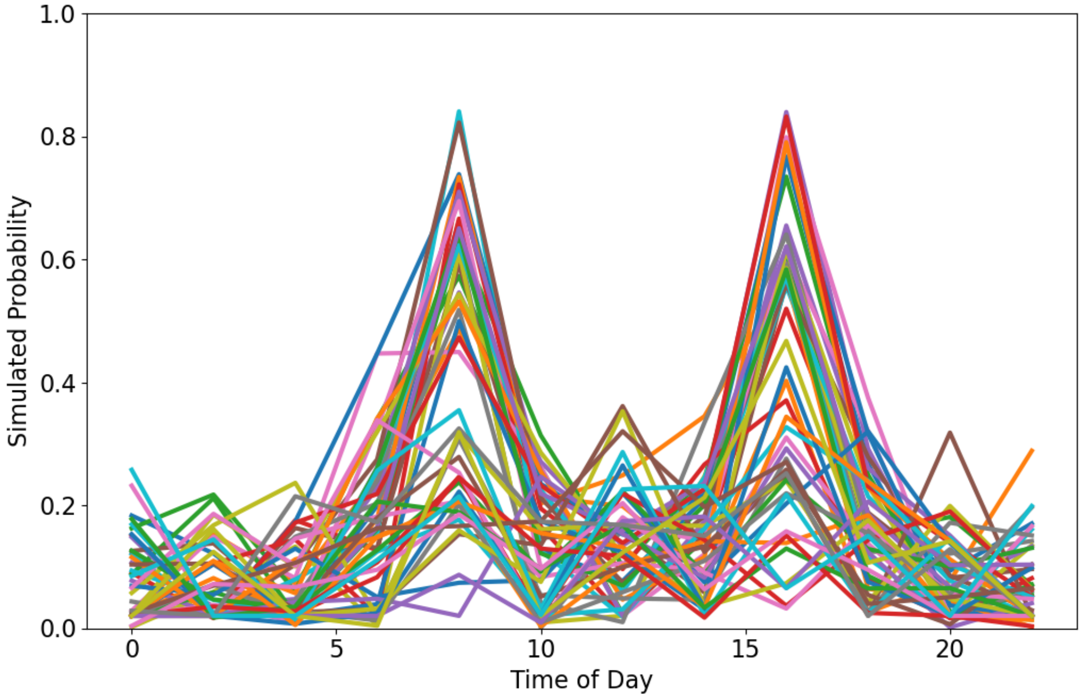


Figure 4.2.3: Example of the population mean for 50 different households simulated using cluster means in Figure 4.2.2.

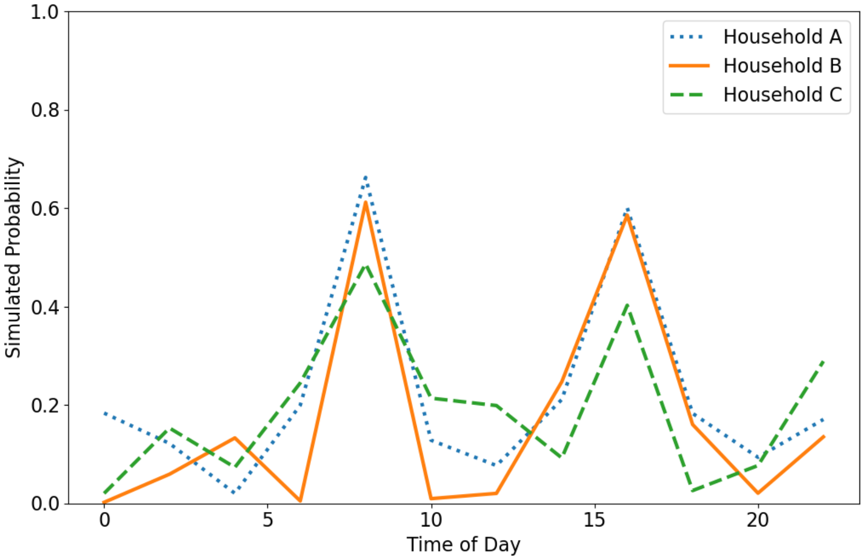


Figure 4.2.5: Simulated probability of Kettle sensor triggering on 'Day 0' for Households A-C, from Figure 4.2.3.

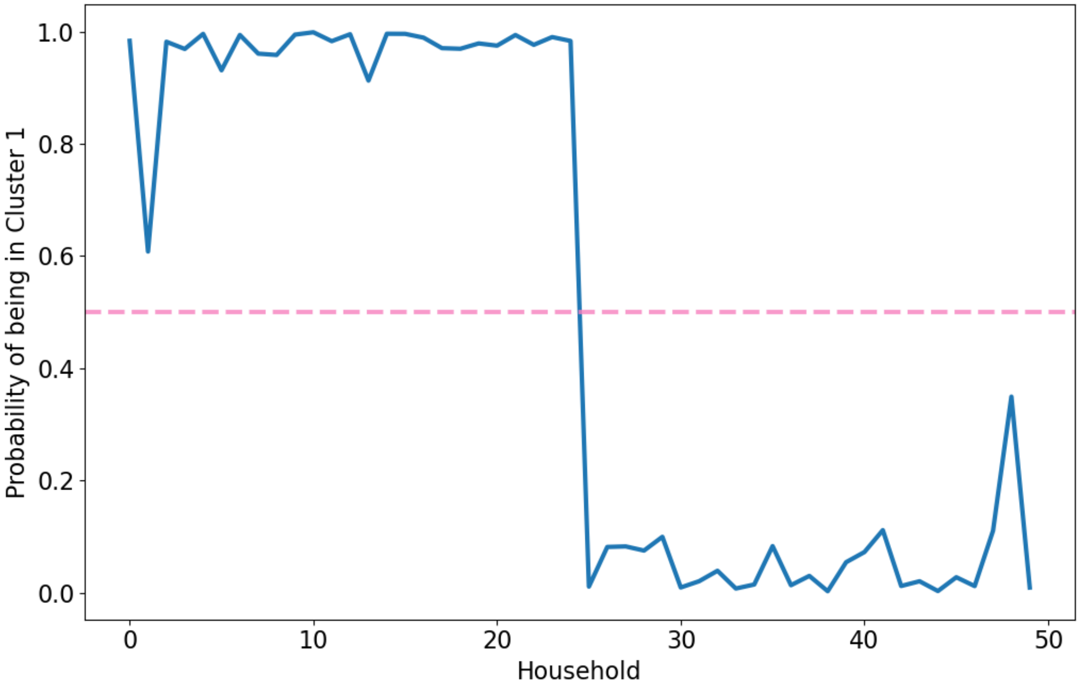


Figure 4.2.4: Probability of being in Cluster 1 for each household on 'Day 1'. The first 25 were simulated from Cluster 1 and the second from Cluster 2.

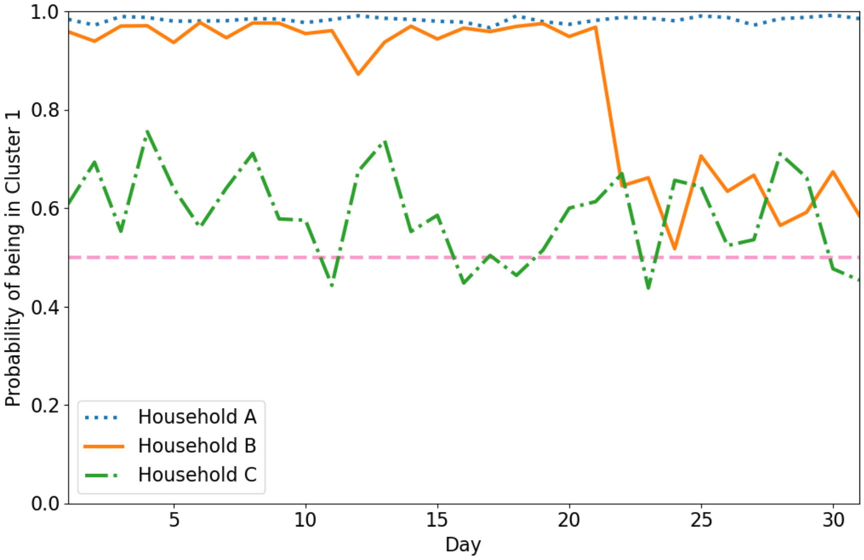


Figure 4.2.6: Probability of being in Cluster 1 for each household across 31 days (pink dashed line indicates the cluster boundary at 0.5).

Figure 4.2.6 shows the cluster membership probability results for Household A to C, presenting the probability of being in Cluster 1 over 31 days. The probability of being in Cluster 1 for Household C is closer to 0.5, as expected from Figure 4.2.5. If we look at this on a household by household basis, we can gather each vector of cluster memberships probabilities into a time series, for day 1 to 31. We can now use this time series to perform cluster analysis.

4.2.3 Changepoint Analysis

Using the cluster membership probabilities for each day from Section 4.2.2, we have a time series for each individual household. We define each column of the probability matrix, Y_i, as the vector of these cluster probabilities, y_t, for time $t \in \{0, \ldots, T\}$ for household i. The matrix can then be defined as follows.

$$Y_i = [y_{i,k,t}] \in (0,1)^{K \times T},$$

where k refers to the cluster $k \in \{1, \ldots, K\}$. For ease of understanding we drop i from within matrix notation for matrix Y_i. This leads to the continuing property that each column for $t \in \{0, \ldots, T\}$,

$$\sum_{k=1}^{K} y_{k,t} = 1. \tag{4.2.1}$$

We assume that each column is a realisation from a Dirichlet distribution as the output has similar properties, i.e. it must follow the property in Equation 4.2.1. The Dirichlet distribution is often used as a conjugate prior for the categorical distribution and multinomial distribution. We define the Dirichlet distribution Probability Density

Function using our notation as,

$$f(y_{t,1},\dots,y_{t,K}) = \frac{1}{\mathrm{B}(\alpha)}\prod_{k=1}^{K} y_{k,t}^{\alpha_h-1}. \tag{4.2.2}$$

To estimate the log-likelihood we use the method from Minka (2000) to estimate the parameters, $\hat{\alpha}_h$, for the Dirichlet distribution. The changepoint method used for this data is the Pruned Exact Linear Time method (PELT) (Killick et al., 2012). PELT is chosen as it is able to adapt to differing log-likelihoods and accurately infer the number of changepoints and their locations within the data.

We assume each column of Y_i is an observation from a Dirichlet distribution with a set of parameters, then at some column, τ, the parameters of the Dirichlet distribution change. To set this up, we let the model have m changepoints with the set of positions $\tau_{1:m} = \{\tau_1,\dots,\tau_m\}$. Let $\tau_0 = 0$ and $\tau_{m+1} = T$. Let $\tau_0 = 0$ and $\tau_{m+1} = T$. The set of changepoint positions is ordered such that $\tau_t < \tau_s$ if and only if $t < s$. For $t \geq s$, the observations from time s to time t are given as $y_{s:t} = \{y_s,\dots,y_t\}$, where y_t is the column of Dirichlet observations. The general framework for a changepoint model is to minimise a cost function. We use the negative log-likelihood of the Dirichlet distribution, $\mathcal{C}((y_{(\tau+1):s}))$, plus a penalty function, $g(T,K)$,

$$\mathcal{C}((y_{(\tau+1):s})) = -\sum_{i=\tau+1}^{s} \log\left(f(y_{t,1},\dots,y_{t,K}|\alpha_1,\dots,\alpha_K)\right).$$

Common choices for the penalty function are the Bayesian Information Criterion (BIC) (Schwarz, 1978) and Modified Bayesian Information Criterion (MBIC) (Zhang and Siegmund, 2007) which are linear in T. We choose to use the BIC penalty to showcase how the method works with simple penalties, however, choosing a penalty function is something that has been studied in detail (Haynes et al., 2017). Alternative penalties may be better suited for different data.

We embed the Dirichlet log-likelihood estimates into the PELT framework. For the simple example we use the three scenarios discussed at the end of Section 4.2.2, Household A with no change in behaviour, B which moves cluster towards the end of the month, and C whose behaviour is closer to the boundary between the two clusters.

Household A finds no changes within the data as expected. Household B detects a change at day 21, which is clear to see in Figure 4.2.6 with the reduced probability of being in the first cluster. The probability for Household B in Figure 4.2.6 is always above 0.5, therefore if we only considered cluster membership this change would not be identifiable. Household C lies close to the boundary between two clusters, i.e. it will occasionally be found in cluster 2, while predominantly being in cluster 1. The changepoint analysis does not find a change, despite the change in cluster membership. This is ideal for our application as while Household C has a varied routine, it is regular behaviour for the household. These examples highlight the benefits of this framework using Dirichlet observations.

Popular clustering techniques can find a change in cluster membership, however the significance of this change is not assessed. Through our implementation of the Dirichlet distribution into changepoint techniques, we can determine whether this is a change in behaviour for the household.

4.2.4 Generalised Framework

For our application, we wish to group households with similar behaviour together to see when they deviate from group behaviour. Therefore, we want to monitor each household over time within this group structure and use this information to decide when this is an indicator for a change in household behaviour. This results in the need for a method that is able to output a result that indicates with what probability a household belongs to each group. Then for each household we use these probabilities to identify when a change in group is something of concern using changepoint methods. This results in the general framework in Algorithm 2. A practitioner is welcome to substitute the techniques we have describe here for identifying changes in cluster probabilities with other techniques provided that the inputs/outputs are able to feed into the next phase of the pipeline.

Algorithm 2 Generalised Framework for Identifying Changes in Probabilistic Clustering Membership

```
Input:
T, the total number of time points in the time series, for $i \in \{0, \ldots, T\}$.
$x_i$, a matrix of items to be clustered, where each row is a different item, e.g. household.
Known cluster means and variances.

Step 1: Get the clusters probabilities.

for $i \in 0 : T$ do
    Cluster $x_i$ using known means, return probability of each point being in each of the clusters.
end for

Note: If updating cluster means, set current time $i$ as time 0 and continue in Step 1.
Step 2: Model cluster probabilities as a Dirichlet distribution.
Step 3: Identify changes in cluster probabilities. Perform changepoint analysis on a each individual household to identify changes in the probabilities of belonging to each cluster.
Return: Vector of changepoint locations, cluster membership probabilities for each item at each time point.
```

4.3 Simulation Study

We now present a simulation study to showcase this framework, using the Howz data as inspiration. First we simulate days from a Dirichlet distribution to show the efficacy of the changepoint analysis when using the estimate log-likelihoods from the Dirichlet. Next we present results running through the entire framework to test how well this method works for a simple edge case, i.e. when a household is in the middle of two clusters. Finally we simulate three clusters from three Howz households that showcase different behaviour Howz would like to differentiate between.

We signify a change in behaviour, if the method is able to identify the correct number of changepoints and the identified changepoints are within a small window of the true change. The simulated data may produce outliers, however, we choose this as a reasonably strict classifier to evaluate our method.

4.3.1 Simulating from the Dirichlet

To show that the framework we have described works sufficiently, we must first test that the adaptation of the PELT method we proposed, using the Dirichlet distribution, is able to correctly identify changepoints.

We simulate 100 days from the Dirichlet distribution for the probability of being within 3 clusters for the following ‘households’. Then run the changepoint analysis discussed in Section 4.2.3 to see if we detect a change. We repeat this experiment 500 times to gather the false positive and true detection rates.

For a simplification of the parameters described in equation 4.2.2, we let $J = 3$

for the clusters and $[\alpha_{j,t}]_{j=1}^{J} = [\alpha_j]_{j=1}^{3}$ for this section. For the first three households, we simulate 100 days from the same $[\alpha_1, \alpha_2, \alpha_3]$ to test if the changepoint analysis is flagging false changes. We then simulate 50 days from one set of alphas, $[\alpha_1, \alpha_2, \alpha_3]$ followed by a further 50 days with a change in behaviour given by the vector, $[\beta_1, \beta_2, \beta_3]$.

We choose a minimum segment length of 3, such that the true detection of a changepoint is defined as finding one changepoint between 47 and 53 (inclusive). The false positives indicate that the method has found a changepoint where it should have found none. We run the changepoint detection method over the full 100 days.

To discuss the results in Table 4.3.1, we define the mean and the variance of the Dirichlet distribution are as follows.

$$\mathbf{E}[Y_h] = \frac{\alpha_j}{\sum_{j=1}^{3} \alpha_j} \tag{4.3.1}$$

$$\mathbf{Var}[Y_h] = \frac{\tilde{\alpha}_j(1 - \tilde{\alpha}_j)}{\alpha_0 + 1} \tag{4.3.2}$$

$$\text{where } \tilde{\alpha}_j = \frac{\alpha_j}{\alpha_0} \text{ and } \alpha_0 = \sum_{j=1}^{3} \alpha_j \tag{4.3.3}$$

Hence from this we can see that mean for A_1 is (0.1,0.8,0.1), and the other households can be defined similarly. For every household apart from A_4, A_8 and A_9, this results in a standard deviation for each Y_h to be between 0.03 and 0.05 approximately. This provides a good amount of variation given the (0,1) restriction. We can see in Table 4.3.1, the first three households show that the false positive rate is controlled well, with A_2 being an example of a boundary household, with small variation. Household A_4, the standard deviation is between 0.12 and 0.15, this will result in the household often changing cluster membership. This is another example of a household on the boundary between clusters with a larger variance that A_2. The false positive is con-

Household	$[\alpha_1, \alpha_2, \alpha_3]$	$[\beta_1, \beta_2, \beta_3]$	False Positive Rate
A_1	(10, 80, 10)		0.030
A_2	(40, 20, 40)		0.026
A_3	(60, 20, 20)		0.032
A_4	(6, 2, 2)		0.038
			True Detection Rate
A_5	(10, 80, 10)	(15, 75, 15)	0.920
A_6	(10, 80, 10)	(20, 60, 20)	0.936
A_7	(10, 80, 10)	(30, 40, 30)	0.942
A_8	(100, 800, 100)	(150, 700, 150)	0.948
A_9	(1, 8, 1)	(1.5, 7, 1.5)	0.304
A_{10}	(10, 80, 10)	(1, 8, 1)	0.920
A_{11}	(40, 20, 40)	(50, 20, 30)	0.936
A_{12}	(40, 20, 40)	(50, 10, 40)	0.940
A_{13}	(40, 20, 40)	(10, 80, 10)	0.954
A_{14}	(60, 20, 20)	(50, 20, 30)	0.940
A_{15}	(60, 20, 20)	(40, 20, 40)	0.946
A_{16}	(60, 20, 20)	(10, 80, 10)	0.970

Table 4.3.1: The false positive and true detection rate from Dirichlet distribution simulations with columns; household number, $[\alpha_j]_{j=1}^3$ probabilities before the change, $[\beta_j]_{j=1}^3$ probabilities before the change and the false positive/true detection rate.

trolled well, as desired. Similarly for the true detection rate, we can measure how effective the changepoint analysis is in correctly identifying the changes in the simulated Dirichlet distribution. We see how the variance of the Dirichlet distribution effects the results. For households A_5, A_8 and A_9, the means of each of the α parameters will be the same, however, we have changed the variance. For A_8, the standard deviation is approximately 0.01 for each α, however it is approximately between 0.09 and 0.14 for A_9. This results in the true detection for A_8 to be higher than A_5 as the change in mean will be more visible to the changepoint method. However, as expected it also results in A_9 struggling to correctly detect the change due to the larger variance resulting in the data having 30% detection rate. We also include household A_{10} as an example of a household whose behaviour becomes more erratic, i.e. the same mean but larger variance, so in reality this might be that we are seeing more behaviour similar to the first or third cluster. This is a clear change in behaviour and the changepoint analysis is able to detect this well as expected.

4.3.2 Test using Gaussian Clusters

One of the benefits the proposed framework tackles, is the identification of the significance of cluster membership. If a household often switches between clusters, whether this is a significant change is difficult to identify if the output is just the cluster name. This is regular behaviour for a household as this is their observed pattern, so we expect the changepoint analysis to echo this and not identify a change.

We have chosen 15 minute intervals throughout a day to represent a household's behaviour, which results in a 96 dimensional vector we wish to cluster. We now provide a simple example to show how the framework identifies if a change has occurred when a household lies between two clusters.

Cluster	Mean	Variance
1	(1,0,. . . ,0)	diag(0.09)
2	(3,0,1,. . . ,1)	diag(0.09)
3	(1.5,. . . ,1.5)	diag(0.09)

Table 4.3.2: Mean and Variance for 3 Simple Clusters.

We have set the mean and variance for the 3 chosen clusters in Table 4.3.2, also setting the weights for the Gaussian Mixture modelling to be equal. The variances are chosen such that we can expect little overlap between the first element of the vector, for data simulated as Cluster 1 or 2. We then place a point, P, in between Cluster 1 and 2 to investigate how this effects the clustering.

$$P = (2, 0, 0.5, \dots, 0.5)$$

We simulate the data for each household using the multivariate normal distribution with differing means and variances in Table 4.3.3, and replicate each scenario 500 times with 100 days for each household. Using the set cluster means and variances, we use the method described in Section 4.2.2, to decide for each of the 100 days which cluster it resides in and output the 3 dimensional vector of cluster probabilities. We then run each of the replications through the changepoint analysis to estimate the false positive and true detection rates.

Similar to Section 4.3.1, we set the minimum segment length to 3, hence changepoints are counted as correctly detecting if one is found between 47 and 53 (inclusive) for C$_6$ and two are found between (27 and 33), (67 and 73) (inclusive) for C$_7$.

Household	Type of Household	False Positive Rate
C$_1$	Mean P with variance, diag(0.0001)	0.006
C$_2$	Mean P with variance, diag(0.01)	0.008
C$_3$	Mean P with variance, diag(0.09)	0.002
C$_4$	Cluster 1	0.006
		True Detection Rate
C$_5$	Mean P with variance, diag(0.0001) then changes to variance, diag(0.01) after day 50	0.970
C$_6$	Mean P with variance, diag(0.01) then changes to variance, diag(0.09) after day 50	0.824
C$_7$	Cluster 1 then changes to cluster 2 after day 50	0.998
C$_8$	Cluster 1 then changes to P after day 30 and then cluster 2 after day 70	0.980

Table 4.3.3: Results for different scenarios with columns; household number, type of scenario for household and false positive/true detection rate .

Table 4.3.3 shows that the framework controls the false positive for the first 4 scenarios. Even when we increase the variance in households C_1 to C_3, this shows that the method correctly identifies that a change has not occurred for households on the boundary between two clusters. Similarly the method is very accurate in detecting the true changepoints in C$_5$. We also include C$_6$, for when a household becomes more erratic switching between two clusters after day 50. We can see for C$_7$, that method is able to pick up the change in behaviour very well. For C$_8$, the method has a reduced detection rate, this is due to the change in variance sometimes not resulting in a change seen in the clustering output, i.e. Gaussian mixture modelling is very certain which cluster the new data point should be in at both variances. Hence, if the user wishes for an increased level of granularity than represented here, we suggest using alternative clustering methods. Overall, these results show the benefits of using clustering and changepoint analysis in conjunction for this application, as we have shown that cases that lie between two clusters can consistently be identified correctly.

4.3.3 Using Households from Howz to Simulate clusters

We present a simulated example using Kettle sensors taken from the Howz data as the basis for potential clusters. Taking 3 Kettle sensors as the mean for the clusters shown in Figure 4.3.1, we consider scenarios with 3 different behaviours; A where we have a spike in probability in the morning and afternoon, B spiking only in the morning and C with low level activity throughout the day. Given that the data is restricted to being between 0 and 1, we set a very small variance for each household. For Household A and B, 0.006 of each time interval, and 0.001 for Household C to ensure that the low level activity is maintained. We also define a new mean, Q, the midpoint of Household A and B.

We then design the scenarios in Table 4.3.4. We simulate 100 days for each household and repeat for 500 simulations. Similar to the previous sections, we set the minimum segment length as 3 and a changepoint is correctly detected for households 7 to 9, if one changepoint is detected and it is between 47 and 53 (inclusive).

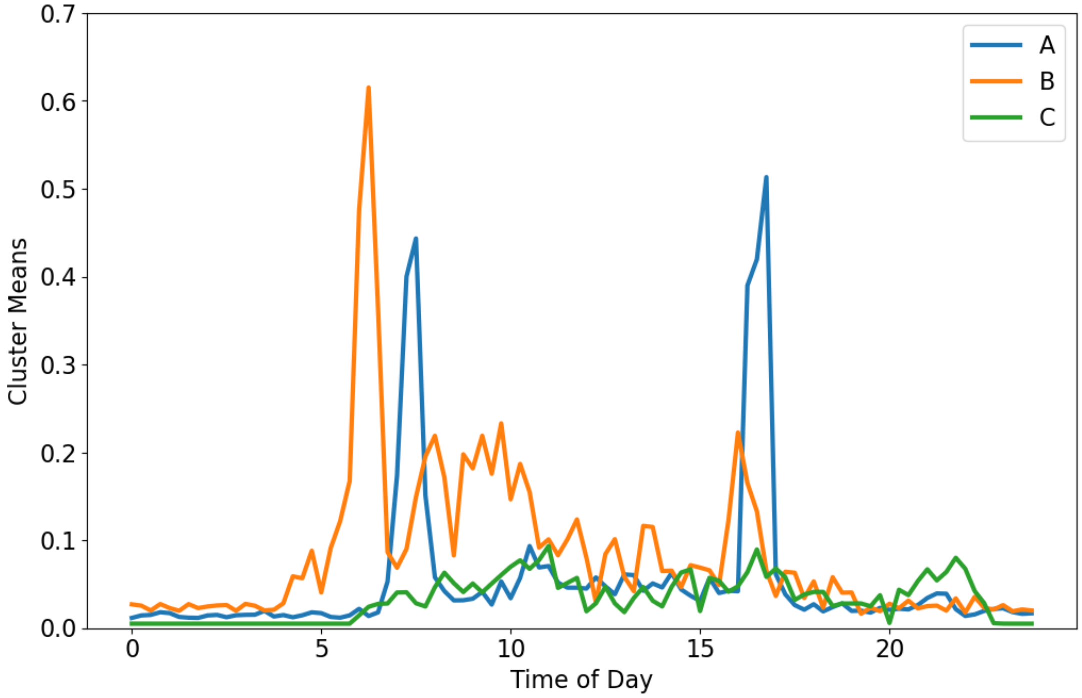


Figure 4.3.1: 3 Kettle sensors from the Howz dataset used as means for the clusters.

Household	Type of Household	False Positive Rate
1	Household A	0.000
2	Household A with larger variance, diag(0.01)	0.024
3	Household B	0.000
4	Household C	0.000
5	Mean Q with variance, diag(0.006)	0.008
		True Detection Rate
6	Household A for the first 50 days then changes to B	0.992
7	Household A for the first 50 days then changes to C	0.998
8	Mean Q with variance, diag(0.006) for 50 days then changes to A	0.988

Table 4.3.4: Results for different scenarios with columns; household number, type of scenario for household and false positive/true detection rate.

Using the clusters defined above, we run the test households through the framework, producing the results in Table 4.3.4. The false positive rate is controlled remarkably well. This shows that the method in this more complex example, is able to correctly identify that household 5, which constantly fluctuates between two clusters, is not exhibiting a change in behaviour. Finally, the framework is able to accurately detect true changes in households 6 to 8. Specifically, it is very encouraging to see that when a household stops fluctuating between two clusters to sit in one in Household 8, this is consistently identified as a change.

4.4 Howz Data Example

We wish to detect if a household moves between three different behaviours when using the Kettle; using the Kettle in the morning, using the Kettle in the morning and afternoon, and low level of activity throughout the day. We acknowledge that any data that characterises the household's behaviour is potentially suitable, however we have chosen this due to ease of access. We provide simple means, as could have been obtained by Howz, for these clusters in Figure 4.4.1.

We have access to 89 households with a Kettle sensor over 8 weeks of data, we run the method from Gillam et al. (2022), to output the 96 dimensional vectors for 6 weeks of data using the first 2 weeks to train the model, showing the probability of a Kettle triggering in that 15 minute time window over the 42 days. The cluster variance for A is set to 0.005, B is 0.005 then 0.001 after 11am and C is 0.0005. For lower levels of activity, seen in cluster B and C, the variance is reduced ensuring that only households which are exhibiting low level activity at those times are classified as such. We present results from three households using the framework with these cluster means and variances.

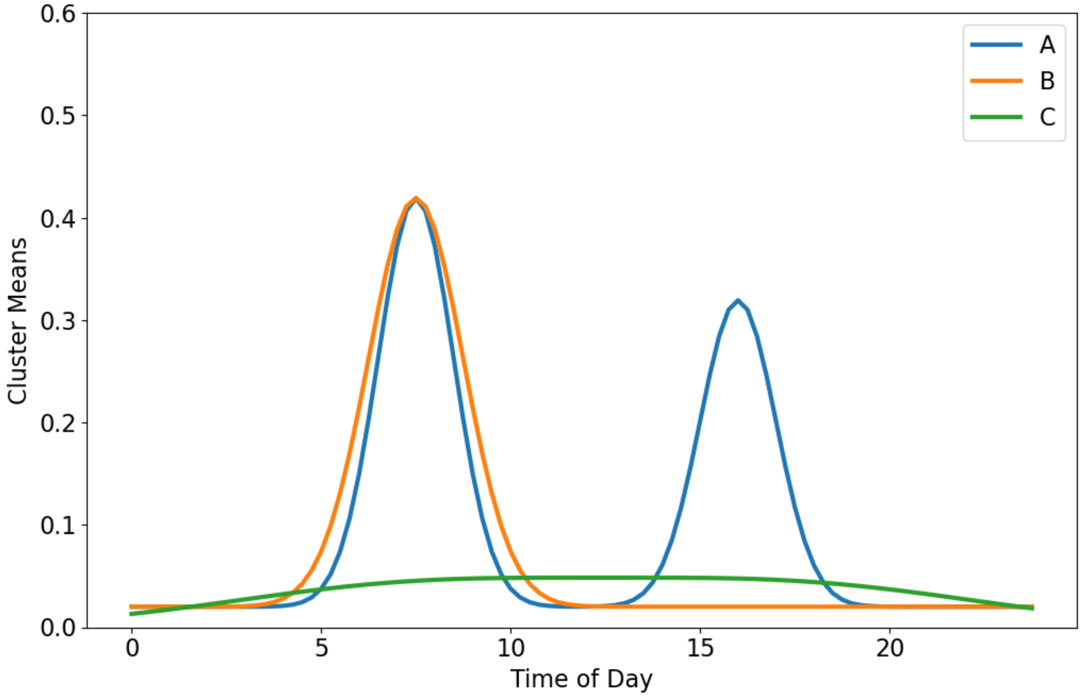


Figure 4.4.1: Cluster means for the Howz Data.

Running each household through the framework, we present the results for 3 different households. Household 1 finds changepoint locations at day 6, 12 and 31. Household 2 and 3 find no changes. To investigate if the framework is working as designed, first we have a look at the output of the cluster membership, taking the largest probability from the output of the Gaussian Mixture Modelling, we have for each day, the most likely cluster. In Figure 4.4.2, we can see that the cluster the data is in changes frequently throughout the 6 weeks. The results of the analysis say that the household begins using the Kettle in the morning frequently, then switches to low level throughout the day, then oscillates between using the Kettle in the morning and using the Kettle in the afternoon. After the last change identified, we see that the household then changes to oscillating between using the Kettle in the morning and low level activity in the day.

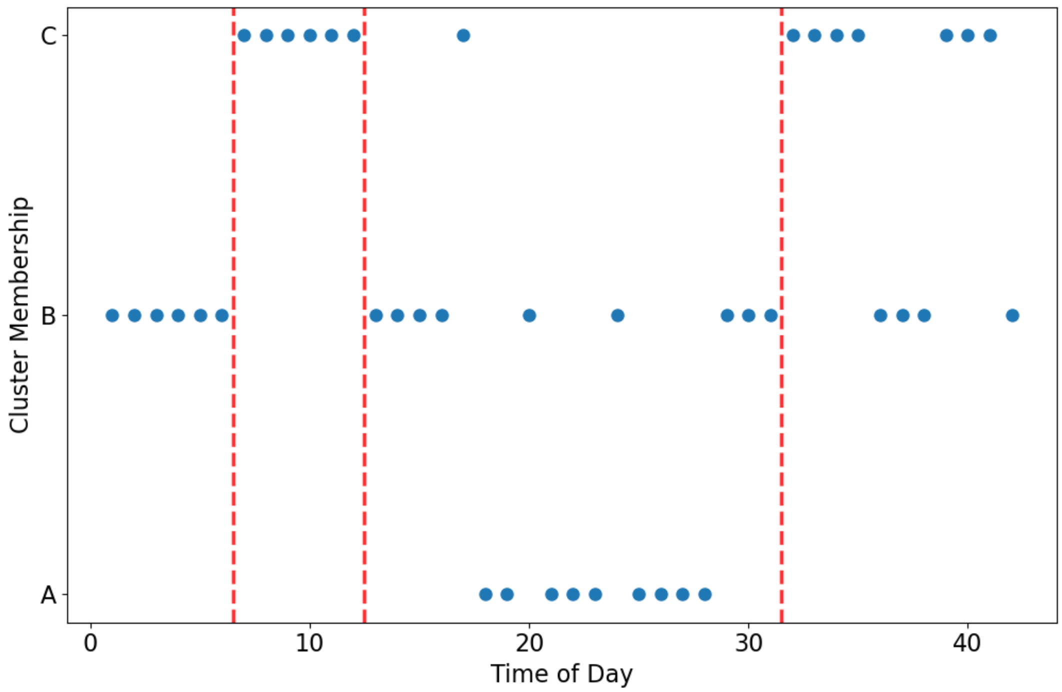


Figure 4.4.2: Cluster membership for Household 1; changepoint locations (red dashed line).

In Figure 4.4.3, we show the average probability over the day for each window between each change found, i.e. between each red dashed line in Figure 4.4.2. These average probabilities confirm what the cluster analysis has shown. For Days 1 to 6, we can see that the household is using the Kettle in the morning and then having low level activity throughout the day. For Days 7 to 12, the spike in the morning has been greatly reduced, resulting in the days looking more like low level activity throughout the day. For days 13 to 31 we can see a large spike in probability in the morning, and a large amount of activity in the day with a small spike in the afternoon, looking more similar to cluster A. Finally, for days 31 to 42 we can see that the spike in the probability is reduced and that it has low level activity throughout the rest of the day. Overall, these averages confirm that the behaviour throughout the 6 weeks is changing as the framework has predicted. Rather than trying to find meaning in the outputted cluster memberships, the changepoint analysis can place emphasis on when the behaviour of the household has changed and is able to handle alternating behaviour for the household in days 13 to 42. This is especially important for this sort of data, as we expect that some households will have a larger amount of daily variation than others.

For Household 2, the cluster membership is consistently cluster B. Whereas the cluster membership is primarily A for Household 3, however, on day 19 the household is believed to be in cluster B. For both of these households, a change in behaviour is not detected. This is echoed in Figure 4.4.4 and 4.4.5, where we can clearly see that Household 2 uses the Kettle primarily in the morning and Household 3 uses the Kettle in the morning and afternoon. This indicates the method is able to accurately place households into the clusters they are most similar to and correctly find no changes in behaviour. Overall, when using households from Howz, the results show very interesting insights into different household's behaviour, that are sensible given the average predicted probabilities.

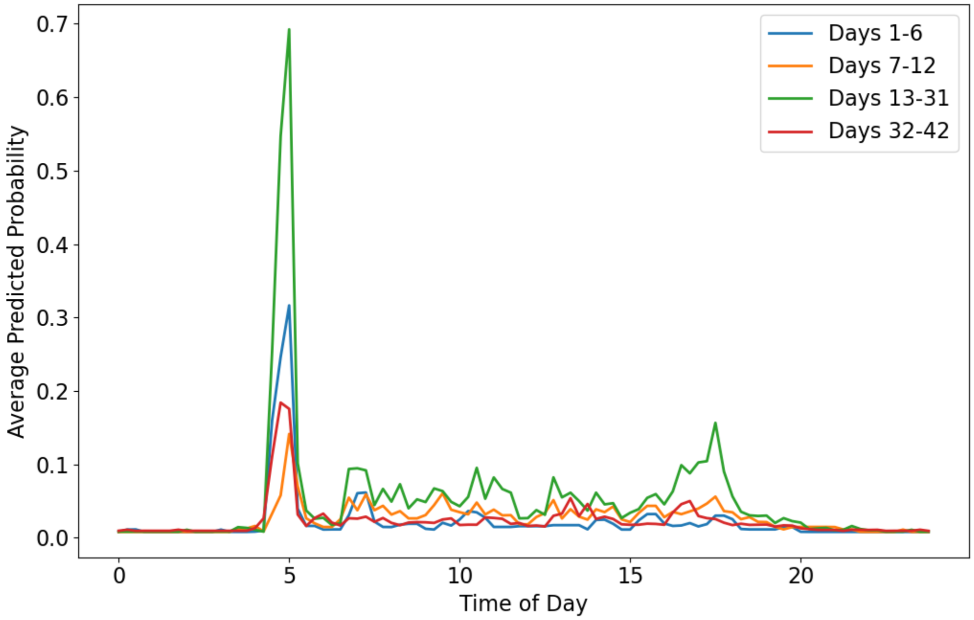


Figure 4.4.3: Average predicted probabilities of the Kettle being triggered for Household 1, for each Window.

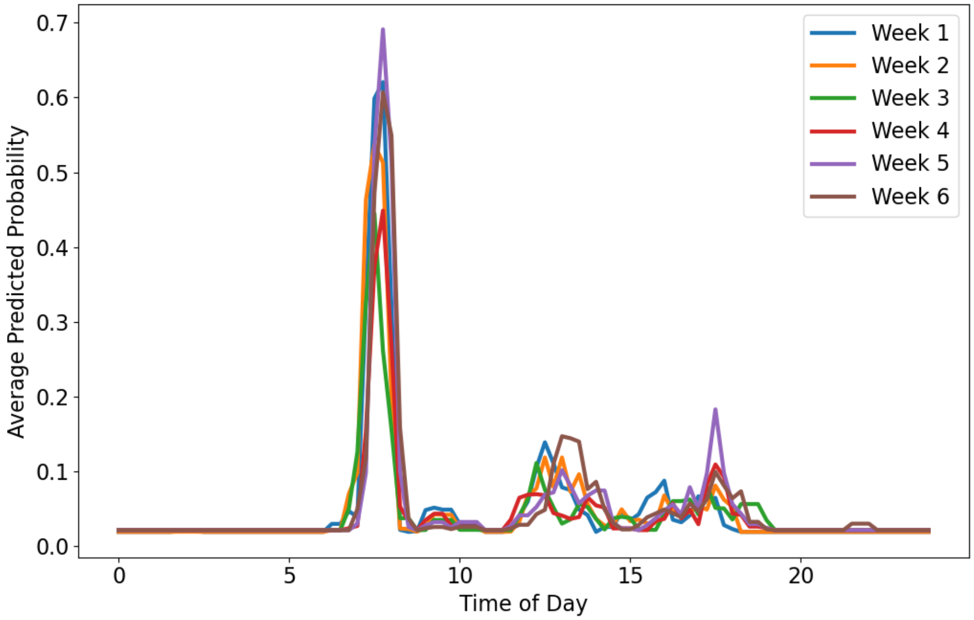


Figure 4.4.4: Average predicted weekly probabilities of a Kettle being triggered for Household 2.

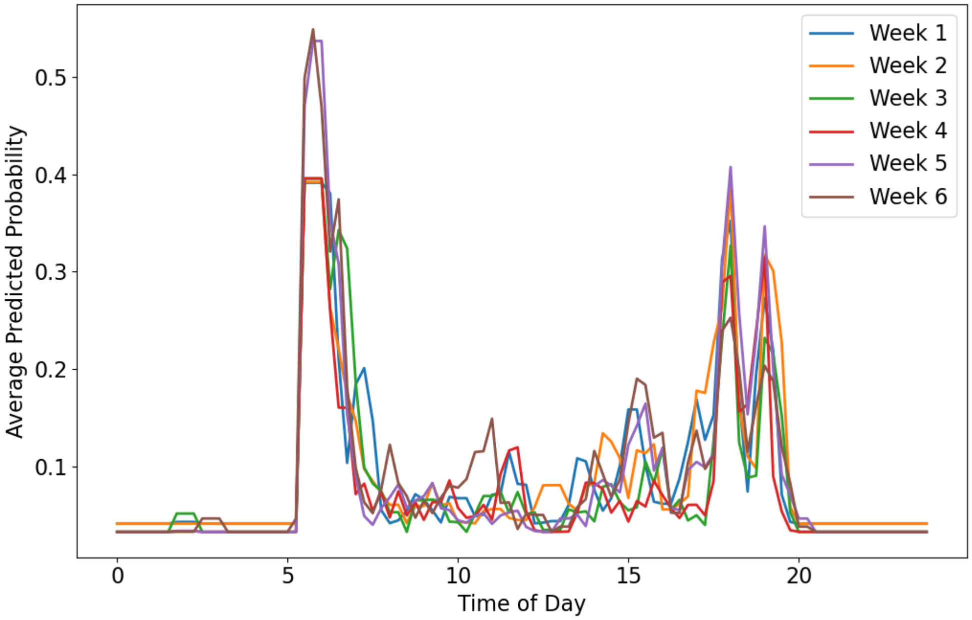


Figure 4.4.5: Average predicted weekly probabilities of a Kettle being triggered for Household 3.

4.5 Conclusion and Future

We have provided a new framework to detect changes in behaviour over time, through monitoring clustering probabilities. We have shown through Section 4.3 and 4.4, that this method is able to accurately detect when a household's behaviour is changing and when it is not. The results presented in Section 4.3, when adapted to use the Dirichlet distribution, works well at correctly identifying changes in simulated behaviour. It is able to accurately detect subtle changes in the probabilities of belonging to each cluster. Specifically, when a household fluctuates on the boundary between two clusters, it is able to detect whether or not a change occurs.

For Section 4.4, we presented a simple cluster mean and variance for three types of behaviour of interest to Howz to demonstrate the insight that could be gained from this framework. A future avenue would be to consider a much larger set of clusters and to use the expert knowledge and further analysis to identify the clusters to link the cluster behaviour to the behaviour seen in the households.

Some households may have very different behaviour day to day. For example, a household may be within one group for weekdays but a different one for weekends, i.e. very active on the weekends with visitors but not very mobile in the week. This is regular behaviour for the household so should be counted as such. However, if this weekend behaviour slowly decreases over time such that they remain in the weekday cluster, we want to detect it as a change. Ushakova et al. (2021) have designed a method to investigate cases such as this, we could implement the Dirichlet probabilities into their approach.

Whilst we have used Gaussian Mixture models to obtain our cluster prediction probabilities these are not constrained to (0,1). We have shown that they work well in practice and allow practitioners to use unconstrained data within the framework. However, we note that for small/large values close to 0/1 the mixture densities will go outside this (0,1) range and thus the cluster probabilities will be incorrect (likely under estimated). A future avenue of exploration could truncate the Gaussian Mixture model or replace with an alternative distribution defined on the (0,1) space.

It is noted by the authors, that the current lack of accumulation of error throughout the framework should be acknowledged. One possible way to tackle this would be using a clustering method which can output error statistics and propagate this error forwards throughout the pipeline. Alternatively inputting different quantiles on the initial dataset could give a boundary for the behaviour of a household.

Chapter 5

Conclusions

5.1 Discussion

This thesis has focused on developing novel methods and frameworks to identify changes in a household's behaviour, using sensor data provided by Howz. We began by performing an exploratory analysis on the data, to provide the reader with understanding of the dataset and intuition for the contributions in chapters 2 to 4.

In Chapter 2 we introduced a novel method for identifying changes in ordered categorical sequences, on a day to day basis. The main idea was to compare a new sequence (day) to previous behaviour for the household, using Pattern Silhouettes which incorporate information from the sub-routines seen within the data to address the first research aim in Section 1.2. We used a new similarity metric, designed to account for the small and varied sequences seen in the data to complete the other research aims. Through simulations and analysis on the Howz dataset, we have shown that the method is able to accurately detect changes in behaviour. In the Howz system, an identification of a change would result in an alert sent to the household, family and carers. Through discussions with the users this can provide context for the change and identify if further steps to ensure the well-being of the household are needed. Within the health care setting, this method can be used on any categorical dataset relating to a patient, able to identify when user-specific changes occur. This is easily adapted to other categorical datasets within home sensor technology. For example; changes in smart meter usage with low, medium, high categories.

Chapter 3 introduced a new model for characterising sensor behaviour and predicting whether a specific sensor will trigger in the next 15 minutes to aid in completing our research aims. The proposed approach is able to follow the peaks and troughs in the sensor data, by taking inspiration from Hawke's processes. We incorporate autoregressive terms to consider the prior data and the other sensors to satisfy the final research aim. We showed, through simulations and use on different households, that the method is able to outperform the Logistic regression model with the same covariates. Within Howz, this method can be used to inform the household about their interactions with the sensors. In practice the quantile figures could be used to show how we expect a user to interact with each sensor over a certain time period e.g., monthly.

Chapter 4, developed a novel framework for identifying changes in behaviour of an individual household by comparing to other households who follow similar behaviours. The framework combines clustering and changepoint techniques to follow a household through time, detecting when they deviate from regular behaviour. This identifies how related households are to each group and identifies the significance when they deviate away, in line with the research aims outlined in Section 1.2. We are also able to identify subtle changes in probabilistic cluster membership when households lie between two groups, that popular clustering methods are not suited for. Through simulations we showed the framework is able to accurately identify changes in behaviour, while allowing for households which are on the boundaries between clusters. Given the generalised framework, this method could be extended to work on all types of devices found within a home health system to monitor symptoms and progression of illnesses. For example in monitoring patients with dementia. As discussed in Section 1, clinicians have knowledge of the different symptoms and are able to differentiate between early and late stages (Alzheimer's Association, 2020). This could be taken as clusters, using the information from various devices (including static sensors and medical devices). Monitoring and identifying changes in households symptoms can then be identified through the framework.

While these methods were designed considering the health management application, they can be easily applied to data from any field with the same structure. For example; migration of birds, if we consider types of birds as categories and the time they have been spotted in migration, we have a categorical dataset. These methods can then provide insight on particular routes of migration to identify changes in migration seasons. The framework in chapter 4 is generalised further, given any data suitable for clustering, as long as the conditions of the framework are met, this method can provide insight on the significance of cluster membership throughout time. Therefore this method could be used on for clustering any data with a time component. For example; for audio or textual data (e.g. changes in speech patterns over time), within cyber security (e.g. changes in clustering of log data) or on measurements taken from medical or environmental sensors (e.g. monitoring of blood pressure in groups of heart disease patients).

5.2 Future Work

We conclude this thesis by considering further ideas to extend the research in each chapter. In chapter 2, we consider a day of data as an ordered categorical sequence, without considering the time aspect explicitly. Incorporating time into the model could provide better identification of changes in behaviour, for routines that only occur at certain times of the day. A simple extension, would be to separate the day into different time windows, e.g. night time, morning, etc. Then running the method for each subsection of data. This will provide more inference on how time of day affects behaviour. Alternatively, we could add a penalty to the similarity function based on the time difference of the silhouettes we are comparing. However, this would provide new challenges to ensure maximal packing constraint holds.

Due to the increasing size of the sensors datasets, we instead chose to consider the time aspect in a different way, resulting in the work presented in chapter 3. We chose to model each sensor individually, as previous work focused on activity as a whole (Taylor et al., 2021). The advantage for Howz is that this allows them to choose which sensors to monitor on a household by household basis; tailoring to the needs of the clients. A further extension would to identify changes in the forecast residuals (Grundy et al., 2022). The challenge here is in designing a threshold; due to the small probabilities, it may prove difficult to identify a threshold sensitive to meaningful changes. We could also model the joint distribution of sensors, as this could provide further inference on the interactions between the sensors.

In chapter 4, we used the output from chapter 3 as an example of one characterisation of the data that could be used to identify a change in behaviour. An important extension would be to account for the accumulation of error throughout the framework, which could be incorporated by outputting error statistics at each step. However, this would depend on the users choice of clustering and changepoint techniques, so further work could look into how certain methods interact with each other. As this method provides a framework, an interesting avenue of research for this application could be decreasing the computational time. One way to accomplish this would be using a different characterisation of the data that is faster to compute.

Appendix A

Identifying Irregular Activity Sequences

A.1 Maximal Packing Explanation

Discussed in Section 2.3.3, there are two different types of silhouette: a silhouette where the start and end are the same $s = s'$ and one where they are different $s \neq s'$.

To consider overlap for the first case, it is clear the maximum number of silhouettes occurs when the sequence is entirely the same event.

Therefore the maximum number of times a SQS can occur is $|X| - k'$.

For the second case, the start and end events are different. The maximal way of packing the silhouettes is to take the SQS of length k and replace the first k' elements inside with the start event. For (D,K,k'):

$$\rlap{$\overbrace{}^{k}$}\underbrace{D\ D\ D\ D\ \dots\ D}_{k'}\ \underbrace{K\ K\ \dots\ K}_{k'}.$$

Hence, k' is the maximum number of ways to fit a SQS into a space of length $2k'$. For a sequence of length $|X|$ divisible by $2k'$, $k'\lfloor\frac{|X|}{2k'}\rfloor$ is the maximum number of sequence silhouette occurrences. However, for sequences that do not divide by $2k'$, the residual part of the sequence has length $v = |X| \mod 2k'$, where the modulo operator returns the remainder of the Euclidean division. If $v < k$ then the residual sequence is too short for for the silhouette, but it is possible to pack additional silhouettes if $k' < v < 2k'$. Hence:

$$w_X = \begin{cases} 0 & v < k \\ v \bmod k' & \text{otherwise.} \end{cases}$$

Therefore, for the different cases, the maximum number of occurrences for a sequence silhouette for one sequence where $|X|$ is the length of the sequence,

$$\text{Maximal Packing} = \begin{cases} |X| & k' = 0 \quad s = s', \\ |X| - k' & k' \geq 1 \quad s = s', \\ k'\left\lfloor\frac{|X|}{2k'}\right\rfloor + w_X & k' \geq 1 \quad s \neq s'. \end{cases}$$

A.2 Maximum Score Explanation

For the first case where the SQS satisfies $s = s'$, we simply have that each sequence silhouette instance within the maximally packed sequence will have a match at each element (see table A.2.1).

Sequence: X	$\underline{\mathrm{D}}$	D	D	D	$\underline{\mathrm{D}}$				(D , D , 4)
Sequence: Y	$\underline{\mathrm{D}}$	*D*	**D**	D	$\underline{\mathrm{D}}$	*D*	**D**	D	Total
Maximum Score	β	λ	2λ	3λ					$\beta + 6\lambda$
		β	λ	2λ	3λ				$\beta + 6\lambda$
			β	λ	2λ	3λ			$\beta + 6\lambda$
				β	λ	2λ	3λ		$\beta + 6\lambda$

Table A.2.1: Example of a maximum score for maximally packed sequence silhouette instance $(D, D, 4)$.

Hence, the score when comparing every instance between two sequences, X and Y, for silhouettes with the same start and end event is:

$$\max(\Gamma(X \mid Y, S) = \begin{cases} \beta \; |X| \; |Y| & \text{if } k' = 0 \text{ and } s = s', \\ (|X| - k') \; (|Y| - k') \left(k'\beta + \frac{k'(k'-1)\lambda}{2}\right) & \text{if } k' \geq 1 \text{ and } s = s', \end{cases}$$

This scoring method does not hold for the second case where the silhouette start and end events are different, $s \neq s'$. In table A.2.2 we see that the score is reduced for each sequential aligned silhouette due to the increased number of mismatches.

To find a general formula for the second case, we consider one maximally packed instance in the new sequence and find the score compared to a maximally packed regular sequence. The results use a counter,

$$\{c \in \mathbb{N}_0 \;:\; 0 \leq c < k', k' = k - 1, k \in \{1, \ldots, \mathcal{K}\}\}.$$

Consider only $k' > 1$ as this is the first chance for overlap to start. For $c = 0$, we

Sequence: X	$\underline{\text{D}}$	D	D	D	$\underline{\text{K}}$				(D , K , 4)	
Sequence: Y	$\underline{\text{D}}$	*D*	**D**	D	$\underline{\text{K}}$	*K*	**K**	K	Total	c
Maximum Score	β	λ	2λ	3λ					$\beta + 6\lambda$	0
		β	λ	2λ	2λ				$\beta + 5\lambda$	1
			β	λ	λ	λ			$\beta + 3\lambda$	2
				β	0	0	0		β	3

Table A.2.2: Example of a maximum score for maximally packed sequence silhouette Instance $(D, K, 4)$. Counter c used in score calculation.

have the first row in table A.2.2. The maximum possible score is $\beta + \sum_{m=1}^{k'-1} m\lambda$, as every element matches. Given the structure of the maximally packed sequence silhouettes, silhouette match $c \in \{1, \ldots, k'-1\}$ will have matches until encountering the first mismatch, and then the remainder will be mismatches. For $c = k'-1$, everything inside the sequence silhouette will be a mismatch so the score will be β. This results in a general function, $M(c)$ for each c:

$$M(c) = \beta + \begin{cases} \sum_{m=1}^{k'-c-1} m\lambda + (k'-c-1)c\lambda & k' > 1,\ 0 \leq c < k'-1, \\ 0 & k' > 1,\ c = k'-1. \end{cases}$$

Taking β out for each silhouette, the maximum score for a length of $2k'$ is:

$$M^\star = k'\beta + \begin{cases} \lambda & \text{if } k' = 2, \\ \underbrace{\sum_{m=1}^{k'-1} m\lambda}_{c=0} + \underbrace{\sum_{c=1}^{k'-2} \left(\sum_{m=1}^{k'-c-1} m\lambda + (k'-c-1)c\lambda \right)}_{c \in \{1,\ldots,k'-2\}} & \text{if } k' > 2, \\ 0 & \text{otherwise.} \end{cases}$$

$$= k'\beta + \begin{cases} \lambda & \text{if } k' = 2, \\ \frac{\lambda k'(k'-1)(k'+1)}{3} & \text{if } k' > 2, \\ 0 & \text{otherwise.} \end{cases}$$

We now have the maximal score for each silhouette, thus, we calculate the overall score for all instances of the sequence silhouette in the two sequences. Therefore, the maximum score when comparing sequences X and Y is:

$$\max(\Gamma(X \mid Y, S) = \begin{cases} \beta \ |X| \ |Y| & \text{if } k' = 0 \text{ and } s = s', \\ (|X| - k') \ (|Y| - k') \left(k'\beta + \frac{k'(k'-1)\lambda}{2} \right) & \text{if } k' \geq 1 \text{ and } s = s', \\ \beta k \lfloor \frac{|X|}{2k'} \rfloor k' \lfloor \frac{|Y|}{2k'} \rfloor & \text{if } k' = 1 \text{ and } s \neq s', \\ (M^\star \left\lfloor \frac{|Y|}{2k'} \right\rfloor + \sum_{c=1}^{w_Y} \mathrm{M}(c))(k' \left\lfloor \frac{|X|}{2k'} \right\rfloor + w_X) + A(w_X, w_Y) & \text{if } k' > 1 \text{ and } s \neq s'. \end{cases}$$

where $A(w_X, w_Y)$ is an extra term defined in Appendix A.3 to account for subsequent score matching.

A.3 Explanation of Extra Term in Maximal Score

The function $A(w_X, w_Y)$ describes the adjustment to the maximum score with the matching between the residual subsequence of X with the perfectly overlapping subsequences of Y. Specifically, we base the calculations against the maximal packing case where all of the intermediate sequence events for the silhouette are identical to the first, and assess what are the differences in the sequence score as we increase the occurrence of the last event from the right. To illustrate, table A.3.1 presents the similarity score for the SQS $(D, K, 6)$ with all matched offsets when the intermediate events are all D (the initial match parameter β is ignored as this is already accounted for elsewhere in the maximum score formula).

Sequence: X		D	D	D	D	D	D	K						(D , K , 6)
Sequence: Y	Line	D	D	D	D	D	D	K	K	K	K	K	K	Total
Maximum Score	1		λ	2λ	3λ	4λ	5λ							15λ
	2			λ	2λ	3λ	4λ	4λ						14λ
	3				λ	2λ	3λ	3λ	3λ					12λ
	4					λ	2λ	2λ	2λ	2λ				9λ
	5						λ	λ	λ	λ	λ			5λ
	6							0	0	0	0	0		0λ

Table A.3.1: Iteration 1 for score calculation of example sequence silhouette $(D, K, 6)$.

Tables A.3.2—A.3.6 presents the similarity score for the same SQS and sequence Y, but we sequentially replacing the right most intermediate event with the last matched event, i.e. K. In the final column of these tables, we note relationship of the scores with those in table 13, which consists of a swap of the first line with the appropriate iteration index line or a difference the score total.

Sequence: X	D	D	D	D	D	K	K						(D , K , 6)	
Sequence: Y	D	D	D	D	D	D	K	K	K	K	K	K	Total	Diff from I1
Maximum Score		λ	2λ	3λ	4λ	4λ							14λ	Line 2
			λ	2λ	3λ	4λ	5λ						15λ	Line 1
				λ	2λ	3λ	3λ	4λ					13λ	$+1\lambda$
					λ	2λ	2λ	2λ	3λ				10λ	$+1\lambda$
						λ	λ	λ	λ	2λ			6λ	$+1\lambda$
							0	0	0	0	λ		λ	$+1\lambda$

Table A.3.2: Iteration 2 for score calculation of example sequence silhouette $(D, K, 6)$, the final column represents the difference in λ between Iteration 1 and Iteration 2.

Sequence: X	D	D	D	D	K	K	K						(D , K , 6)	
Sequence: Y	D	D	D	D	D	D	K	K	K	K	K	K	Total	Diff from I1
Maximum Score		λ	2λ	3λ	3λ	3λ							12λ	Line 3
			λ	2λ	3λ	3λ	4λ						13λ	-1λ
				λ	2λ	3λ	4λ	5λ					15λ	Line 1
					λ	2λ	2λ	3λ	4λ				12λ	(+1+2)λ
						λ	λ	λ	2λ	3λ			8λ	(+1+2)λ
							0	0	0	λ	2λ		3λ	(+1+2)λ

Table A.3.3: Iteration 3 for score calculation of example sequence silhouette $(D, K, 6)$, the final column represents the difference in λ between Iteration 1 and Iteration 3.

Sequence: X	D	D	D	K	K	K	K						(D , K , 6)	
Sequence: Y	D	D	D	D	D	D	K	K	K	K	K	K	Total	Diff from I1
Maximum Score		λ	2λ	2λ	2λ	2λ							9λ	Line 4
			λ	2λ	2λ	2λ	3λ						10λ	-4λ
				λ	2λ	2λ	3λ	4λ					12λ	0λ
					λ	2λ	3λ	4λ	5λ				15λ	Line 1
						λ	λ	2λ	3λ	4λ			11λ	(+1+2+3)λ
							0	0	λ	2λ	3λ		6λ	(+1+2+3)λ

Table A.3.4: Iteration 4 for score calculation of example sequence silhouette $(D, K, 6)$, the final column represents the difference in λ between Iteration 1 and Iteration 4.

Sequence: X	D	D	K	K	K	K	K						(D , K , 6)	
Sequence: Y	D	D	D	D	D	D	K	K	K	K	K	K	Total	Diff from I1
Maximum Score		λ	λ	λ	λ	λ							5λ	Line 5
			λ	λ	λ	λ	2λ						6λ	-8λ
				λ	λ	λ	2λ	3λ					8λ	-4λ
					λ	λ	2λ	3λ	4λ				11λ	+2λ
						λ	2λ	3λ	4λ	5λ			15λ	Line 1
							0	λ	2λ	3λ	4λ		10λ	(+1+2+3+4)λ

Table A.3.5: Iteration 5 for score calculation of example sequence silhouette $(D, K, 6)$, the final column represents the difference in λ between Iteration 1 and Iteration 5.

Sequence: X	D	K	K	K	K	K	K						(D , K , 6)	
Sequence: Y	D	D	D	D	D	D	K	K	K	K	K	K	Total	Diff from I1
Maximum Score		0	0	0	0	0							0λ	Line 6
			0	0	0	0	λ						λ	-13λ
				0	0	0	λ	2λ					3λ	-9λ
					0	0	λ	2λ	3λ				6λ	-3λ
						0	λ	2λ	3λ	4λ			10λ	+5λ
							λ	2λ	3λ	4λ	5λ		15λ	Line 1

Table A.3.6: Iteration 6 for score calculation of example sequence silhouette $(D, K, 6)$, the final column represents the difference in λ between Iteration 1 and Iteration 6.

These differences are collated into the following matrix that presents all score differences from an exhaustive investigation up to longest SQS considered in this paper ($\mathcal{K} = 7$, see Appendix A.5).

$$F = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & -4 & -8 & -13 \\ 0 & 1 & 0 & 0 & -4 & -9 \\ 0 & 1 & 0 & 0 & 2 & -3 \\ 0 & 1 & 3 & 6 & 0 & 5 \\ 0 & 1 & 3 & 6 & 10 & 0 \end{bmatrix}$$

The values differences seen in the table for the i^{th} iteration correspond to the i^{th} column of F, up to the j^{th} row associated to the SQS length up to k'. This matrix has the property that the values sum to zero. This corresponds to the case where the lengths of sequences X and Y are both divisible by 2k, and so there are no residual subsequence after accounting for optimal packing. The adjustment term in the maximal score is then the sum of all entries in the upper-left $w_Y \times w_X$ submatrix of F, specifically:

$$A(w_X, w_Y) = \sum_{i=1}^{w_X} \sum_{j=1}^{w_Y} F_{j,i}.$$

We note that the entries in matrix F can be expressed formulaically according to:

$$F_{j,i} = \begin{cases} j(j-1) - \frac{i(i-1)}{2} & \text{if } 1 < j < i, \\ \frac{i(i-1)}{2} & \text{if } j > i, \\ 0 & \text{otherwise.} \end{cases}$$

A.4 Tables for Section 2.4

Regular Sequence Probabilities	New Sequence Probabilities	$\lvert R\rvert$	Minimum Sequence Length		
			4	8	10
$(\frac{1}{3},\frac{1}{3},\frac{1}{3})$	$(\frac{1}{3},\frac{1}{3},\frac{1}{3})$	56	0.19	0.29	0
	$(0.5, 0.4, 0.1)$		1	1	1
	$(0.8, 0.1, 0.1)$		1	1	1
$(\frac{1}{3},\frac{1}{3},\frac{1}{3})$	$(\frac{1}{3},\frac{1}{3},\frac{1}{3})$	28	0.02	0.04	0
	$(0.5, 0.4, 0.1)$		1	1	1
	$(0.8, 0.1, 0.1)$		1	1	1
$(\frac{1}{3},\frac{1}{3},\frac{1}{3})$	$(\frac{1}{3},\frac{1}{3},\frac{1}{3})$	14	0.11	0.13	0.08
	$(0.5, 0.4, 0.1)$		1	1	1
	$(0.8, 0.1, 0.1)$		1	1	1
$(0.8, 0.1, 0.1)$	$(0.8, 0.1, 0.1)$	56	0.11	0.13	0
	$(0.7, 0.15, 0.15)$		0.44	0.49	1
	$(0.6, 0.2, 0.2)$		0.90	1	1
	$(\frac{1}{3},\frac{1}{3},\frac{1}{3})$		1	1	1
$(0.8, 0.1, 0.1)$	$(0.8, 0.1, 0.1)$	28	0.13	0.01	0
	$(0.7, 0.15, 0.15)$		0.46	0.37	1
	$(0.6, 0.2, 0.2)$		0.91	1	1
	$(\frac{1}{3},\frac{1}{3},\frac{1}{3})$		1	1	1

Table A.4.1: Each column represents (from left to right); the probabilities the regular sequences are simulated with, the probabilities the new sequences are simulated with, number of regular sequences and the different minimum lengths of the sequences. The elements of the table contain the proportion which reject at least one hypothesis. Maximum sequence length 10.

Regular Sequence Probabilities	New Sequence Probabilities	$\|R\|$	Minimum Sequence Length		
			4	8	10
$(0.8, 0.1, 0.1)$	$(0.8, 0.1, 0.1)$	14	0.08	0.04	0.02
	$(0.7, 0.15, 0.15)$		0.34	0.40	1
	$(0.6, 0.2, 0.2)$		0.79	1	1
	$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$		1	1	1
$(0.65, 0.25, 0.1)$	$(0.65, 0.25, 0.1)$	56	0.13	0.07	0.31
	$(0.8, 0.1, 0.1)$		0.81	1	1
	$(1/3, 1/3, 1/3)$		1	1	1
$(0.65, 0.25, 0.1)$	$(0.65, 0.25, 0.1)$	28	0.16	0.03	0.07
	$(0.8, 0.1, 0.1)$		0.76	1	1
	$(1/3, 1/3, 1/3)$		1	1	1
$(0.65, 0.25, 0.1)$	$(0.65, 0.25, 0.1)$	14	0.13	0.17	0.12
	$(0.8, 0.1, 0.1)$		0.77	1	1
	$(1/3, 1/3, 1/3)$		1	1	1

Table A.4.2: Each column represents (from left to right); the probabilities the regular sequences are simulated with, the probabilities the new sequences are simulated with, number of regular sequences and the different minimum lengths of the sequences. The elements of the table contain the proportion which reject at least one hypothesis. Maximum sequence length 10.

Regular Sequence Probabilities	New Sequence Probabilities	$\lvert R\rvert$	Minimum Sequence Length		
			4	10	50
$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$	$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$	56	0.03	0.07	0.17
	$(0.5, 0.4, 0.1)$		1	1	1
	$(0.8, 0.1, 0.1)$		1	1	1
$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$	$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$	28	0.09	0.35	0.31
	$(0.5, 0.4, 0.1)$		1	1	1
	$(0.8, 0.1, 0.1)$		1	1	1
$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$	$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$	14	0.28	0.34	0.53
	$(0.5, 0.4, 0.1)$		1	1	1
	$(0.8, 0.1, 0.1)$		1	1	1
$(0.8, 0.1, 0.1)$	$(0.8, 0.1, 0.1)$	56	0.05	0.11	0.07
	$(0.7, 0.15, 0.15)$		0.46	0.96	1
	$(0.6, 0.2, 0.2)$		0.98	1	1
	$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$		1	1	1
$(0.8, 0.1, 0.1)$	$(0.8, 0.1, 0.1)$	28	0.10	0.19	0.15
	$(0.7, 0.15, 0.15)$		0.68	1	1
	$(0.6, 0.2, 0.2)$		0.99	1	1
	$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$		1	1	1

Table A.4.3: Each column represents (from left to right); the probabilities the regular sequences are simulated with, the probabilities the new sequences are simulated with, number of regular sequences and the different minimum lengths of the sequences. The elements of the table contain the proportion which reject at least one hypothesis. Maximum sequence length 50.

Regular Sequence Probabilities	New Sequence Probabilities	$\|R\|$	Minimum Sequence Length		
			4	10	50
$(0.8, 0.1, 0.1)$	$(0.8, 0.1, 0.1)$	14	0.21	0.21	0.16
	$(0.7, 0.15, 0.15)$		0.83	0.96	1
	$(0.6, 0.2, 0.2)$		0.98	1	1
	$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$		1	1	1
$(0.65, 0.25, 0.1)$	$(0.65, 0.25, 0.1)$	56	0.07	0.09	0.12
	$(0.8, 0.1, 0.1)$		0.99	1	1
	$(1/3, 1/3, 1/3)$		1	1	1
$(0.65, 0.25, 0.1)$	$(0.65, 0.25, 0.1)$	28	0.14	0.16	0.17
	$(0.8, 0.1, 0.1)$		0.99	1	1
	$(1/3, 1/3, 1/3)$		1	1	1
$(0.65, 0.25, 0.1)$	$(0.65, 0.25, 0.1)$	14	0.25	0.22	0.19
	$(0.8, 0.1, 0.1)$		0.99	1	1
	$(1/3, 1/3, 1/3)$		1	1	1

Table A.4.4: Each column represents (from left to right); the probabilities the regular sequences are simulated with, the probabilities the new sequences are simulated with, number of regular sequences and the different minimum lengths of the sequences. The elements of the table contain the proportion which reject at least one hypothesis. Maximum sequence length 50.

A.5 Extra Simulated Results for Households with No Routines for Different Parameters

A.5.1 Varying the length of $\mathcal{K}$

We now take 56 days as the set of regular sequences and sequence lengths of minimum 4 and maximum 25, from table 2.4.2. We vary $\mathcal{K} = (3, 5, 7)$ to see how it affects the results, including $\mathcal{K} = 3$ for ease of comparison. Set $\beta = 1$ and $\lambda = 0.5$, Table A.5.1 shows little difference in the false detection results for varying $\mathcal{K}$ in this example. Keeping $\mathcal{K}$ small increases the computational efficiency of the method, so this should be considered in the choice of $\mathcal{K}$.

A.5.2 Varying β and λ

We take 56 days as the set of regular sequences and sequence lengths of minimum 4 and maximum 25, from table 2.4.2. We vary β and λ, including $\beta = 1$ and $\lambda = 0.5$ for ease of comparison. Table A.5.2 shows little difference overall across the varying parameter values. As you can see the false positive rate and the small change shows a slight decrease as λ increases. However, for the larger change, $(0.6, 0.2, 0.2)$, we can see it increases very slightly and then decreases. As the difference between the results is so minimal, we choose a suitable value in the middle for the results ($\beta = 1$ and $\lambda = 0.5$).

Regular Sequence Probabilities	New Sequence Probabilities		$\mathcal{K}$		
		$\|R\|$	3	5	7
$(\frac{1}{3},\frac{1}{3},\frac{1}{3})$	$(\frac{1}{3},\frac{1}{3},\frac{1}{3})$	56	0.038	0.038	0.038
	$(0.5, 0.4, 0.1)$		0.918	0.918	0.918
	$(0.8, 0.1, 0.1)$		0.930	0.930	0.930
$(0.8, 0.1, 0.1)$	$(0.8, 0.1, 0.1)$	56	0.022	0.028	0.040
	$(0.7, 0.15, 0.15)$		0.236	0.258	0.330
	$(0.6, 0.2, 0.2)$		0.928	0.928	0.922
	$(\frac{1}{3},\frac{1}{3},\frac{1}{3})$		1	1	1
$(0.65, 0.25, 0.1)$	$(0.65, 0.25, 0.1)$	56	0.036	0.036	0.042
	$(0.8, 0.1, 0.1)$		0.846	0.838	0.834
	$(\frac{1}{3},\frac{1}{3},\frac{1}{3})$		0.994	0.994	0.994

Table A.5.1: Changing k for varying sequence lengths between 4 and 25. Each column represents (from left to right); the probabilities the regular sequences are simulated with, the probabilities the new sequences are simulated with, number of regular sequences and the different minimum lengths of the sequences. The elements of the table contain the proportion which reject at least one hypothesis.

A.5.3 Increasing the Number of Sensors

As above, we take 56 days as the set of regular sequences and varying sequence lengths of minimum 4 and maximum 25. We now explore having 4 sensors to see how it affects

Regular Sequence Probabilities	New Sequence Probabilities		(β, λ)				
		$\|R\|$	(1,0.1)	(1,0.25)	(1,0.5)	(1,0.75)	(1,0.9)
$(0.8, 0.1, 0.1)$	$(0.8, 0.1, 0.1)$	56	0.024	0.022	0.022	0.022	0.022
	$(0.7, 0.15, 0.15)$		0.248	0.236	0.236	0.230	0.230
	$(0.6, 0.2, 0.2)$		0.924	0.926	0.928	0.922	0.922
	$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$		1	1	1	1	1

Table A.5.2: Changing β and λ for varying sequence lengths between 4 and 25. Each column represents (from left to right); the probabilities the regular sequences are simulated with, the probabilities the new sequences are simulated with, number of regular sequences and the different β and λ. The elements of the table contain the proportion which reject at least one hypothesis.

the results. Set $\beta = 1$ and $\lambda = 0.05$ to ensure the maximal packing constraint holds. The results in the table are similar to the three sensor case. However, the method struggles to pick up the smaller changes. This is due to the small sequence lengths limiting the number of patterns that can be observed. This increases the variability within the reference set which reduces the power of the method to identify small changes.

Regular Sequence Probabilities	New Sequence Probabilities	$\lvert R\rvert$	Minimum Sequence Length 4	10	25
$(\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4})$	$(\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4})$	56	0.044	0.040	0.000
	$(0.5, 0.3, 0.1, 0.1)$		0.860	1	1
	$(0.7, 0.1, 0.1, 0.1)$		0.882	1	1
$(0.7, 0.1, 0.1, 0.1)$	$(0.7, 0.1, 0.1, 0.1)$	56	0.010	0.026	0.014
	$(0.6, 0.15, 0.15, 0.1)$		0.076	0.378	1
	$(0.5, 0.2, 0.2, 0.1)$		0.502	0.988	1
	$(\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4})$		0.998	1	1
$(0.6, 0.3, 0.05, 0.05)$	$(0.6, 0.3, 0.05, 0.05)$	56	0.026	0.040	0.044
	$(0.7, 0.1, 0.1, 0.1)$		0.906	1	0.914
	$(\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4})$		0.998	1	0.996

Table A.5.3: 4 sensors for varying sequence lengths between 4 and 25. Each column represents (from left to right); the probabilities the regular sequences are simulated with, the probabilities the new sequences are simulated with, number of regular sequences and the different lengths of $\mathcal{K}$. The elements of the table contain the proportion which reject at least one hypothesis.

Appendix B

Modelling and Forecasting of at Home Activity

B.1 Simplifying the Link Function

Assuming $b_0 = 0$ and $y_0 = 0$ to initialise we can simplify b_t to;

$$\begin{aligned}
b_t &= \phi_b b_{t-1} + \pi_b y_{t-1} \\
&= \phi_b \left[\phi_b b_{t-2} + \pi_b y_{t-2}\right] + \pi_b y_{t-1} \\
&= \phi_b^t b_0 + \left[\phi_b^{t-1} \pi_b y_{b,0} + \phi_b^{t-2} \pi_b y_1 + \cdots + \phi_b \pi_b y_{t-2} + \pi_b y_{t-1}\right] \\
&= \pi_b \left[\phi_b^{t-2} y_1 + \cdots + \phi_b y_{t-2} + y_{t-1}\right] \\
&= \pi_b \sum_{i=1}^{t-1} \phi_b^{i-1} y_{t-i}.
\end{aligned}$$

Similarly, assuming $c_{j,0} = 0$ and $z_{j,0} = 0$,

$$c_{j,t} = \tau_j \sum_{i=1}^{t-1} \psi_j^{i-1} z_{j,t-i}. \tag{B.1.1}$$

If we set $y_{-1} = 0$, $y_0 = 0$ and consider $t \geq 96$,

$$\begin{aligned}
d_t &= \phi_d d_{t-96} + \pi_d \max(y_{t-96-1:t-96+1}) \\
&= \phi_d \left[\phi_d d_{t-96(2)} + \pi_d \max(y_{t-96*2-1:t-96*2+1})\right] + \pi_b \max(y_{t-96-1:t-96+1}) \\
&= \pi_d \Bigg[\phi_d^{\lfloor \frac{t}{96} \rfloor} \max(y_{t-96\lfloor \frac{t}{96} \rfloor-1:t-96\lfloor \frac{t}{96} \rfloor+1}) + \dots \\
&\qquad + \phi_d \max(y_{t-96*2-1:t-96*2+1}) + \max(y_{t-96-1:t-96+1})\Bigg] \\
&= \pi_d \sum_{i=1}^{\lfloor \frac{t}{96} \rfloor} \phi_d^{i-1} \max(y_{t-96i-1:t-96i+1}) \qquad \text{for } t \geq 96.
\end{aligned}$$

Therefore, Δ_t and p_t can be expressed as,

$$\begin{aligned}
\Delta_t &= a + \pi_b \sum_{i=1}^{t-1} \phi_b^{i-1} y_{t-i} + \sum_{j=1}^{J} \tau_j \sum_{i=1}^{t-1} \psi_j^{i-1} z_{j,t-i} \\
&\quad + \begin{cases} \pi_d \sum_{i=1}^{\lfloor \frac{t}{96} \rfloor} \phi_d^{i-1} \max(y_{t-96i-1:t-96i+1}) & \text{if } t \geq 96, \\ 0 & \text{Otherwise.} \end{cases}, \\
p_t &= \frac{\exp(\Delta_t)}{1 + \exp(\Delta_t)}.
\end{aligned}$$

B.2 Derivation of Log-Likelihood

Given the parameter set, $\alpha = \{a, \pi_b, \phi_b, \tau_1, \psi_1, ..., \tau_j, \psi_j, \pi_d, \phi_d, \}$, with $\{y_t\}$ from a Bernoulli distribution, we can express the Likelihood as,

$$\begin{aligned} L\left(\alpha|\{y_t\}_{t=1}^{T}\right) &= \prod_{t=1}^{T} p_t^{y_t}\left(1-p_t\right)^{1-y_t} \\ &= \prod_{t=1}^{T}\left(\frac{\exp(\Delta_t)}{1+\exp(\Delta_t)}\right)^{y_t} \times \left(1-\frac{\exp(\Delta_t)}{1+\exp(\Delta_t)}\right)^{1-y_t} \\ &= \prod_{t=1}^{T}\frac{\exp(y_t\Delta_t)}{\left(1+\exp(\Delta_t)\right)^{y_t}\left(1+\exp(\Delta_t)\right)^{1-y_t}} \\ &= \prod_{t=1}^{T}\frac{\exp(y_t\Delta_t)}{1+\exp(\Delta_t)}. \end{aligned}$$

We now consider the log-likelihood in the estimation of α, to find

$$\begin{aligned} l(\alpha|\{y_t\}_{t=1}^{T}) &= \log\left(\prod_{t=1}^{T}\frac{\exp(y_t\Delta_t)}{1+\exp(\Delta_t)}\right) \\ &= \sum_{t=1}^{T}\log\left(\frac{\exp(y_t\Delta_t)}{1+\exp(\Delta_t)}\right) \\ &= \sum_{t=1}^{T} y_t\Delta_t \; - \; \sum_{t=1}^{T}\log\left(1+\exp(\Delta_t)\right). \end{aligned}$$

B.3 Gradient of the Log-Likelihood

We define the gradient of the log-likelihood as

$$\nabla l\left(\alpha\right) = \left\{\frac{\partial l\left(\alpha\right)}{\partial a}, \frac{\partial l\left(\alpha\right)}{\partial \pi_b}, \frac{\partial l\left(\alpha\right)}{\partial \phi_b}, \frac{\partial l\left(\alpha\right)}{\partial \tau_1}, \frac{\partial l\left(\alpha\right)}{\partial \psi_1}, \ldots, \frac{\partial l\left(\alpha\right)}{\partial \tau_j}, \frac{\partial l\left(\alpha\right)}{\partial \psi_j}, \frac{\partial l\left(\alpha\right)}{\partial \pi_d}, \frac{\partial l\left(\alpha\right)}{\partial \phi_d}\right\}$$

Assuming $y_{-1} = 0$, $y_0 = 0$, and $z_{j,0} = 0$ for $j = 1 \ldots, J$. We can find the partial derivatives of the log-likelihood as

$$\begin{aligned}
\frac{\partial l(\alpha)}{\partial a} &= \sum_{t=1}^{T} y_t - \sum_{t=1}^{T} \frac{\exp(\Delta_t)}{1+\exp(\Delta_t)} \\
\frac{\partial l(\alpha)}{\partial \pi_b} &= \sum_{t=1}^{T} y_t \sum_{i=1}^{t-1} \phi_b^{i-1} y_{t-i} - \sum_{t=1}^{T} \frac{\exp(\Delta_t)}{1+\exp(\Delta_t)} \sum_{i=1}^{t-1} \phi_b^{i-1} y_{t-i} \\
\frac{\partial l(\alpha)}{\partial \phi_b} &= \pi_b \sum_{t=1}^{T} y_t \sum_{i=2}^{t-1} (i-1)\, \phi_b^{i-2} y_{t-i} - \pi_b \sum_{t=1}^{T} \frac{\exp(\Delta_t)}{1+\exp(\Delta_t)} \sum_{i=2}^{t-1} (i-1)\, \phi_b^{i-2} y_{t-i}.
\end{aligned}$$

For other sensors $j = 1, \ldots, J$:

$$\begin{aligned}
\frac{\partial l(\alpha)}{\partial \tau_j} &= \sum_{t=1}^{T} y_t \sum_{i=1}^{t-1} \psi_j^{i-1} z_{j,t-i} - \sum_{t=1}^{T} \frac{\exp(\Delta_t)}{1+\exp(\Delta_t)} \sum_{i=1}^{t-1} \psi_j^{i-1} z_{j,t-i} \\
\frac{\partial l(\alpha)}{\partial \psi_j} &= \tau_j \sum_{t=1}^{T} y_t \sum_{i=2}^{t-1} (i-1)\, \psi_j^{i-2} z_{j,t-i} - \tau_j \sum_{t=1}^{T} \frac{\exp(\Delta_t)}{1+\exp(\Delta_t)} \sum_{i=2}^{t-1} (i-1)\, \psi_j^{i-2} z_{j,t-i}.
\end{aligned}$$

For the seasonal component with $t \geq 96$:

$$\begin{aligned}
\frac{\partial l(\alpha)}{\partial \pi_d} &= \sum_{t=1}^{T} y_t \sum_{i=1}^{\left\lfloor \frac{t}{96} \right\rfloor} \phi_d^{i-1} \max\left(y_{t-96i-1:t-96i+1}\right) \\
&\quad - \sum_{t=1}^{T} \frac{\exp(\Delta_t)}{1+\exp(\Delta_t)} \sum_{i=1}^{\left\lfloor \frac{t}{96} \right\rfloor} \phi_d^{i-1} \max\left(y_{t-96i-1:t-96i+1}\right). \\
\frac{\partial l(\alpha)}{\partial \phi_d} &= \pi_d \sum_{t=1}^{T} y_t \sum_{i=2}^{\left\lfloor \frac{t}{96} \right\rfloor} (i-1)\, \phi_d^{i-2} \max\left(y_{t-96i-1:t-96i+1}\right) \\
&\quad - \pi_d \sum_{t=1}^{T} \frac{\exp(\Delta_t)}{1+\exp(\Delta_t)} \sum_{i=2}^{\left\lfloor \frac{t}{96} \right\rfloor} (i-1)\, \phi_d^{i-2} \max\left(y_{t-96i-1:t-96i+1}\right).
\end{aligned}$$

B.4 Howz Data Examples

B.4.1 Household 1

Taking a different sensor from this household, Fridge Door, the model selection process identifies the Hallway, Kettle and Lounge sensors as informative but not the Fridge Door sensor itself. The estimated parameters are given in Table B.4.1 and the prediction interval in Figure B.4.1.

α	$\tau_{Hallway}$	$\psi_{Hallway}$	τ_{Kettle}	ψ_{Kettle}	τ_{Lounge}
-3.634	0.848	-0.446	1.027	0.084	0.680
ψ_{Lounge}	π_d	ϕ_d			
0.209	0.392	0.809			

Table B.4.1: Parameter estimates from fitting the model in Section 3.2 to Household 1, Fridge Door sensor.

Again the method follows the change in number of events well with 2 events outside the interval. For this sensor, we see the household uses the Fridge in the morning, around lunch time and then in the evening with low level usage throughout the day.

B.4.2 Household 2

Household 2 has 4 sensors; Bedroom, Hallway, Kettle and Microwave. In Section 3.3.2 we present the Bedroom sensor. We now give the other sensors in Household 2, starting with the Kettle sensor, which after performing the model selection chooses the Hallway and Bedroom but not the Kettle sensor itself. The estimates of the

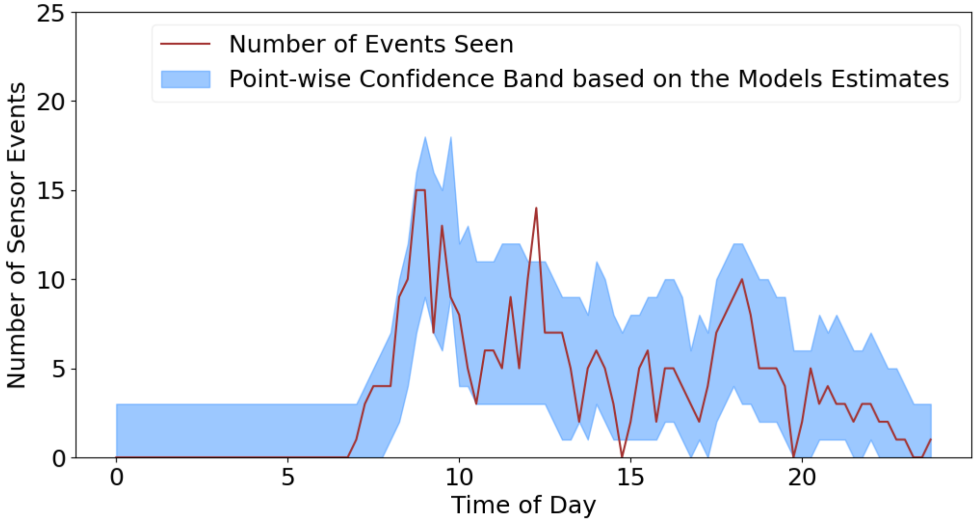


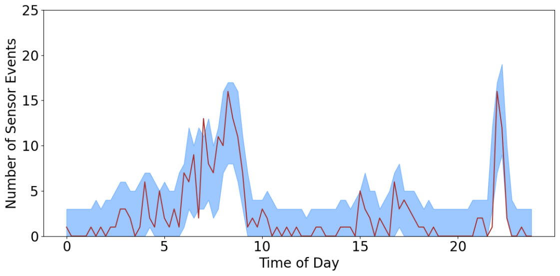
Figure B.4.1: Number of events seen over 27 days (red) with the predicted 95% quantile interval (blue) for the Fridge Door Sensor, Household 1.

parameters are given in Table B.4.2. The estimation for the 95% quantile interval for February is given by Figure B.2(b) which shows the household has 3 clear time periods throughout the day the kettle is used; morning, lunch and afternoon. The method follows the peaks well even with the reduced amount of data due to the kettle being used infrequently.

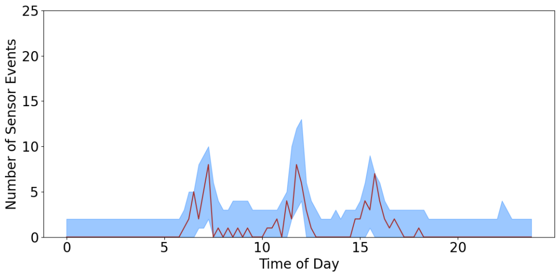
α	$\tau_{Hallway}$	$\psi_{Hallway}$	$\tau_{Bedroom}$	$\psi_{Bedroom}$	π_d	ϕ_d
-4.093	0.807	-0.191	0.765	0.044	0.530	0.871

Table B.4.2: Parameter estimates from fitting the model in Section 3.2 to Household 2, Kettle sensor.

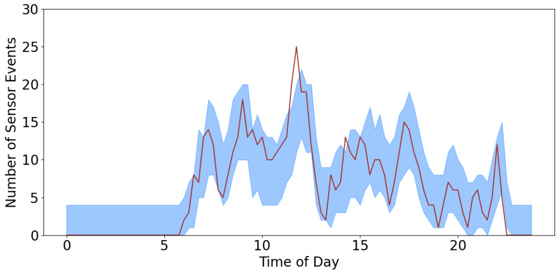
Next we give the Hallway sensor. The model selections chooses the Kettle, Bedroom, autoregressive term and seasonal component. The estimates of the parameters



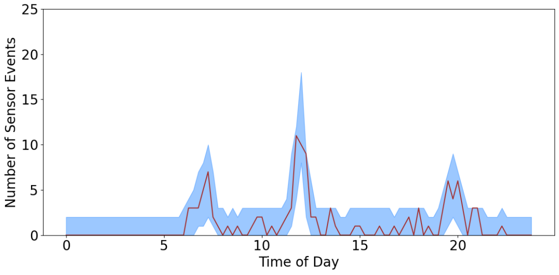
(a) Bedroom Sensor



(b) Kettle Sensor



(c) Hallway Sensor



(d) Microwave Sensor

Figure B.4.2: Number of events seen over 27 days (line) with the predicted 95% quantile interval (shaded band) for each sensor. Please note that these were modelled individually.

are given in Table B.4.3. The estimation for the 95% quantile interval for February is given by Figure B.2(c) which shows the household has behaviour throughout the day. The method follows the peaks well, however, struggles a little to capture the midday spike.

α	π_b	ϕ_b	τ_{Kettle}	ψ_{Kettle}	$\tau_{Bedroom}$
-3.135	1.306	-0.207	0.653	0.933	0.623
$\psi_{Bedroom}$	π_d	ϕ_d			
0.365	0.447	0.798			

Table B.4.3: Parameter estimates from fitting the model in Section 3.2 to Household 1, Hallway sensor.

Finally we give the Microwave sensor, the model selection chooses the Kettle, Hallway and seasonal term (Table B.4.4 shows the parameter estimates). The estimation for the 95% quantile interval for February is given by Figure B.2(d) which shows the household has 3 clear Microwave usage times throughout the day; morning, lunch and evening. Similar to the Kettle sensor, the method follows the peaks well with the reduced amount of data.

α	τ_{Kettle}	ψ_{Kettle}	$\tau_{Hallway}$	$\psi_{Hallway}$	π_d	ϕ_d
-4.263	1.057	-0.705	0.889	-0.244	0.836	0.801

Table B.4.4: Parameter estimates from fitting the model in Section 3.2 to Household 2, Microwave sensor.

The spike parameters, π and τ, give a good indication of the interactions between the sensors. For example in table B.4.1, we see the Kettle spike parameter is negative, so we are unlikely to see the Bedroom sensor if we have just seen the Kettle. However, in table B.4.2, we can see that if we see have seen the bedroom sensor recently we are likely to see the Kettle. The Kettle and Hallway sensors behave differently, looking at the spike parameters in table B.4.2 and B.4.3, we can see that seeing the Kettle recently means we are likely to see the Hallway and vice-versa. The model selection for the Microwave chose the Kettle and Hallway sensor, with both having positive spike parameters, suggesting seeing the Kettle and/or Hallway will mean we are more likely to see the Microwave. However, neither of those chose the Microwave sensor during model selection. While these differing relationships will be unique to the household, it is interesting to see the effect of users interactions with the sensors placed throughout the house.

B.4.3 Household 3

We present a household which has two sensors; Front Door and Kettle. This household has little information, so is a good test for how the method works on the opposite end of the spectrum from Household 1.

We present the results from the Kettle sensor. The model selection process selected the kettle itself and the seasonal component. The parameter estimates are given in Table B.4.5.

Figure B.4.3 shows that even when the sensor is infrequently used and there is little information from the rest of the house, the method is still able to follow the

α	π_b	ϕ_b	π_d	ϕ_d
-2.990	0.159	-0.795	0.322	0.882

Table B.4.5: Parameter estimates from fitting the model in Section 3.2 to Household 3, Kettle sensor.

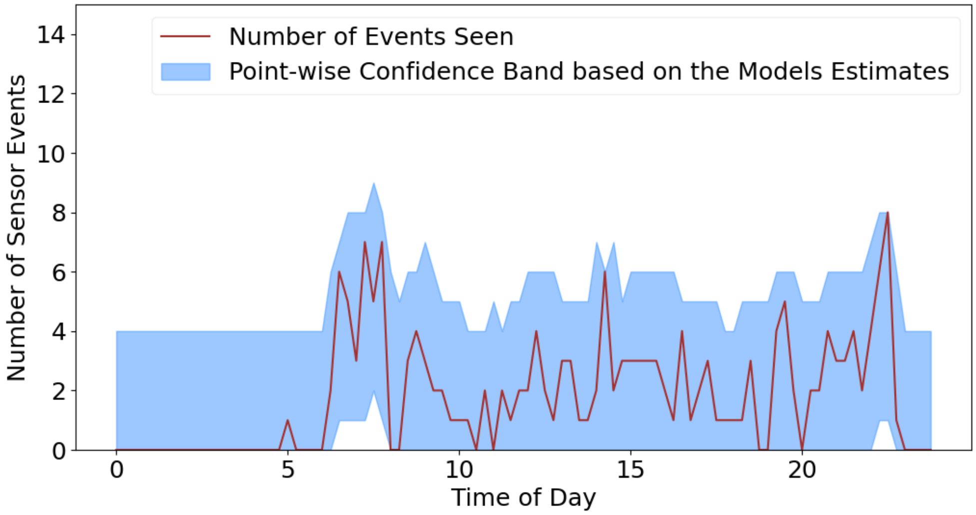


Figure B.4.3: Number of events seen over 27 days (line) with the predicted 95% quantile interval (shaded band) for the Kettle Sensor.

number of true events seen well. Specifically, even with small peaks, the quantile band shifts upwards.

Bibliography

Age UK. Later life in the united kingdom, 2019. Available at `https://www.ageuk.org.uk/`, Accessed : 2019-06-25.

Alzheimer's Association. 2020 alzheimer's disease facts and figures. 2020.

Eder Angelo Milani, Marcelo Hartmann, Marinho G Andrade, and Carlos Alberto Ribeiro Diniz. A bernoulli autoregressive moving average model applied to rainfall occurrence. *Communications in Statistics-Simulation and Computation*, 48(9): 2743–2756, 2019.

Dale J Barr. Analyzing 'visual world'eyetracking data using multilevel logistic regression. *Journal of memory and language*, 59(4):457–474, 2008.

Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300, 1995.

Gessé Dafé, Adriano Veloso, Mohammed Zaki, and Wagner Meira. Learning sequential classifiers from long and noisy discrete-event sequences efficiently. *Data mining and knowledge discovery*, 29(6):1685–1708, 2015.

Richard A Davis, Scott H Holan, Robert Lund, and Nalini Ravishanker. *Handbook of discrete-valued time series*. CRC Press, 2016.

Richard A. Davis, Konstantinos Fokianos, Scott H. Holan, Harry Joe, James Livsey, Robert Lund, Vladas Pipiras, and Nalini Ravishanker. Count time series: A methodological review. *Journal of the American Statistical Association*, 116(535): 1533–1547, 2021.

Anthony Christopher Davison and David Victor Hinkley. *Bootstrap methods and their application*, volume 1. Cambridge university press, 1997.

Norman R Draper and Harry Smith. *Applied regression analysis*, volume 326. John Wiley & Sons, 1998.

Kathryn Dreyer, Adam Steventon, Rebecca Fisher, and Sarah R Deeny. The association between living alone and health care utilisation in older adults: a retrospective cohort study of electronic health records from a london general practice. *BMC geriatrics*, 18(1):269, 2018.

Idris A Eckley, Paul Fearnhead, and Rebecca Killick. Analysis of changepoint models. *Bayesian time series models*, pages 205–224, 2011.

Matthew Engelhard, Hongteng Xu, Lawrence Carin, Jason A Oliver, Matthew Hallyburton, and F Joseph McClernon. Predicting smoking events with a time-varying semi-parametric hawkes process model. *Proceedings of machine learning research*, 85:312, 2018.

Roger Fletcher. *Practical methods of optimization*. John Wiley & Sons, 1987.

Konstantinos Fokianos and Theodoros Moysiadis. Binary time series models driven by a latent process. *Econometrics and Statistics*, 2:117 – 130, 2017.

Philippe Fournier-Viger, Jerry Chun-Wei Lin, Rage Uday Kiran, Yun Sing Koh, and Rincy Thomas. A survey of sequential pattern mining. *Data Science and Pattern Recognition*, 1(1):54–77, 2017.

Dmitriy Fradkin and Fabian Mörchen. Mining sequential patterns for classification. *Knowledge and Information Systems*, 45(3):731–749, 2015.

F. Ge, Y. Ju, Z. Qi, and Y. Lin. Parameter estimation of a gaussian mixture model for wind power forecast error by riemann l-bfgs optimization. *IEEE Access*, 6: 38892–38899, 2018.

Jess Gillam, Rebecca Killick, Jack Heal, and Ben Norwood. Modelling and forecasting of at home activity in older adults using passive sensor technology. *Statistics In Medicine*, (In Press), 2022.

Jess Gillam, Rebecca Killick, and Ben Norwood. Framework to identify subtle changes in probabilistic cluster membership: an application to passive household monitoring. *Submitted*, 2022+a.

Jess Gillam, Rebecca Killick, Simon AC Taylor, Jack Heal, and Ben Norwood. Identifying irregular activity sequences: an application to passive household monitoring. *Submitted*, 2022+b.

Thomas Grundy, Rebecca Killick, and Ivan Svetunkov. Identifying sequential changes in mean and variance within more complex structures. *Submitted*, 2022.

Kaylea Haynes, Idris A Eckley, and Paul Fearnhead. Computationally efficient changepoint detection for a range of penalties. *Journal of Computational and Graphical Statistics*, 26(1):134–143, 2017.

Zengyou He, Simeng Zhang, and Jun Wu. Significance-based discriminative sequential pattern mining. *Expert Systems with Applications*, 122:54–64, 2019.

Michael Höhle. Online change-point detection in categorical time series. In *Statistical modelling and regression structures*, pages 377–397. Springer, 2010.

Yili Hong. On computing the distribution function for the poisson binomial distribution. *Computational Statistics & Data Analysis*, 59:41 – 51, 2013.

Lawrence Hubert and Phipps Arabie. Comparing partitions. *Journal of classification*, 2(1):193–218, 1985.

Anil K Jain, M Narasimha Murty, and Patrick J Flynn. Data clustering: a review. *ACM computing surveys (CSUR)*, 31(3):264–323, 1999.

Sean W Jewell, Toby Dylan Hocking, Paul Fearnhead, and Daniela M Witten. Fast nonconvex deconvolution of calcium imaging data. *Biostatistics*, 2019.

Shwetambara Kekade, Chung-Ho Hseieh, Md Mohaimenul Islam, Suleman Atique, Abdulwahed Mohammed Khalfan, Yu-Chuan Li, and Shabbir Syed Abdul. The usefulness and actual use of wearable devices among the elderly population. *Computer methods and programs in biomedicine*, 153:137–159, 2018.

Yasser Khan, Aminy E Ostfeld, Claire M Lochner, Adrien Pierre, and Ana C Arias.

Monitoring of vital signs with flexible and wearable medical devices. *Advanced materials*, 28(22):4373–4395, 2016.

Rebecca Killick, Paul Fearnhead, and Idris A Eckley. Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, 107(500):1590–1598, 2012.

Tomi Kinnunen, Ilja Sidoroff, Marko Tuononen, and Pasi Fränti. Comparison of clustering methods: a case study of text-independent speaker modeling. *Pattern Recognition Letters*, 32(13):1604–1617, 2011.

Trupti M Kodinariya and Prashant R Makwana. Review on determining number of cluster in k-means clustering. *International Journal*, 1(6):90–95, 2013.

KSS, Office for Life Sciences, and NHS. Technology integated health management : health economic evaluation, 2021. Available at `https://www.howz.com/report-download`, Accessed : 2022-05-30.

Ahmad Lotfi, Amir Pourabdollah, and Diego Resende Faria. Special issue on human behaviour monitoring, interpretation and understanding, 2020.

Geoffrey J McLachlan and Kaye E Basford. *Mixture models: inference and applications to clustering*, volume 38. 1988.

Eder Milani, Marcelo Hartmann, Marinho Andrade, and Carlos Diniz. A bernoulli autoregressive moving average model applied to rainfall occurrence. *Communications in Statistics - Simulation and Computation*, 48:1–17, 12 2018.

Thomas Minka. Estimating a dirichlet distribution, 2000.

Theodoros Moysiadis and Konstantinos Fokianos. On binary and categorical time series models with feedback. *Journal of Multivariate Analysis*, 131:209 – 228, 2014.

Office for National Statistics ONS. National population projections: 2016-based projections, methodology, 2017. Available at `https://www.ons.gov.uk/`, Accessed : 2020-10-13.

Office for National Statistics ONS. Changing trends in mortality: a cross-uk comparison, 1981 to 2016, 2018. Available at `https://www.ons.gov.uk/`, Accessed : 2020-10-13.

Office for National Statistics ONS. Overview of the uk population: january 2021, 2021. Available at `https://www.ons.gov.uk/`, Accessed : 2022-02-09.

Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. On the difficulty of training recurrent neural networks. In *Proceedings of the 30th International Conference on Machine Learning*, pages 1310–1318, 2013.

Joshua Plasse and Niall M Adams. Multiple changepoint detection in categorical data streams. *Statistics and Computing*, pages 1–17, 2019.

Stephen J Preece, John Y Goulermas, Laurence PJ Kenney, Dave Howard, Kenneth Meijer, and Robin Crompton. Activity identification using body-mounted sensors—a review of classification techniques. *Physiological measurement*, 30(4):R1, 2009.

Martin Prince, Martin Knapp, Maelenn Guerchet, Paul McCrone, Matthew Prina, A Comas-Herrera, Raphael Wittenberg, Bayo Adelaja, Bo Hu, Derek King, et al. Dementia uk: Second edition - overview. 2014.

Jun Qi, Po Yang, Atif Waraich, Zhikun Deng, Youbing Zhao, and Yun Yang. Examining sensor-based physical activity recognition and monitoring for healthcare using internet of things: A systematic review. *Journal of biomedical informatics*, 87:138–153, 2018.

R Core Team. *R: A Language and Environment for Statistical Computing.* R foundation for statistical computing, 2020. URL `https://www.R-project.org/`.

Jeffrey S Racine and James G MacKinnon. Inference via kernel smoothing of bootstrap p values. *Computational Statistics & Data Analysis*, 51(12):5949–5957, 2007.

William M Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association*, 66(336):846–850, 1971.

Mayra Z Rodriguez, Cesar H Comin, Dalcimar Casanova, Odemir M Bruno, Diego R Amancio, Luciano da F Costa, and Francisco A Rodrigues. Clustering algorithms: a comparative approach. *PloS one*, 14(1):1–34, 2019.

Gideon Schwarz. Estimating the dimension of a model. *The annals of statistics*, pages 461–464, 1978.

Skipper Seabold and Josef Perktold. statsmodels: econometric and statistical modeling with python. In *9th Python in Science Conference*, 2010.

Simon AC Taylor, Rebecca Killick, Jonathan Burr, and Louise Rogerson. Assessing daily patterns using home activity sensors and within period changepoint detection. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 70(3):579–595, 2021.

Soe Ye Yint Tun, Samaneh Madanian, and Farhaan Mirza. Internet of things (iot) applications for elderly care: a reflective review. *Aging clinical and experimental research*, 33(4):855–867, 2021.

Anastasia Ushakova, Rebecca Killick, and Simon AC Taylor. Multi-level changepoint inference for periodic data sequences. *Submitted*, 2021.

Yasmin van Kasteren, Dana Bradford, Qing Zhang, Mohan Karunanithi, Hang Ding, et al. Understanding smart home sensor data for ageing in place through everyday household routines: a mixed method case study. *JMIR mHealth and uHealth*, 5: e52, 2017.

Y. Wan, Z. Wang, P. Wang, C. Zhang, S. Duan, and N. Li. A comparative study of inversion optimization algorithms for underground metal target detection. *IEEE Access*, 8:126401–126413, 2020.

Yuan H Wang. On the number of successes in independent trials. *Statistica Sinica*, pages 295–312, 1993.

Zhengzheng Xing, Jian Pei, and Eamonn Keogh. A brief survey on sequence classification. *ACM Sigkdd Explorations Newsletter*, 12(1):40–48, 2010.

Hongliang Zhang, Rong Li, Zhennao Cai, Zhiyang Gu, Ali Asghar Heidari, Mingjing Wang, Huiling Chen, and Mayun Chen. Advanced orthogonal moth flame optimization with broyden–fletcher–goldfarb–shanno algorithm: Framework and real-world problems. *Expert Systems with Applications*, 159, 2020.

Nancy R Zhang and David O Siegmund. A modified bayes information criterion with applications to the analysis of comparative genomic hybridization data. *Biometrics*, 63(1):22–32, 2007.