Recognising Activities at Home: Digital and Human Sensors

Jie Jiang
University of Surrey
United Kingdom
jie.jiang@surrey.ac.uk

Riccardo Pozza
University of Surrey
United Kingdom
r.pozza@surrey.ac.uk

Kristrún Gunnarsdóttir
University of Surrey
United Kingdom
k.gunnarsdottir@surrey.ac.uk

Nigel Gilbert
University of Surrey
United Kingdom
n.gilbert@surrey.ac.uk

Klaus Moessner
University of Surrey
United Kingdom
k.moessner@surrey.ac.uk

ABSTRACT

What activities take place at home? When do they occur, for how long do they last and who is involved? Asking such questions is important in social research on households, e.g., to study energy-related practices, assisted living arrangements and various aspects of family and home life. Common ways of seeking the answers rest on self-reporting which is provoked by researchers (interviews, questionnaires, surveys) or non-provoked (time use diaries). Longitudinal observations are also common, but all of these methods are expensive and time-consuming for both the participants and the researchers. The advances of digital sensors may provide an alternative. For example, temperature, humidity and light sensors report on the physical environment where activities occur, while energy monitors report information on the electrical devices that are used to assist the activities. Using sensor-generated data for the purposes of activity recognition is potentially a very powerful means to study activities at home. However, how can we quantify the agreement between what we detect in sensor-generated data and what we know from self-reported data, especially non-provoked data? To give a partial answer, we conduct a trial in a household in which we collect data from a suite of sensors, as well as from a time use diary completed by one of the two occupants. For activity recognition using sensor-generated data, we investigate the application of mean shift clustering and change points detection for constructing features that are used to train a Hidden Markov Model. Furthermore, we propose a method for agreement evaluation between the activities detected in the sensor data and that reported by the participants based on the Levenshtein distance. Finally, we analyse the use of different features for recognising different types of activities.

CCS CONCEPTS

•**Applied computing** → **Sociology;** •**Human-centered computing** → *Collaborative and social computing;* •**Computing methodologies** → *Machine learning;*

KEYWORDS

Sensors, Time use diaries, Activity recognition, Time series, Internet of things, Social research

ACM Reference format:
Jie Jiang, Riccardo Pozza, Kristrún Gunnarsdóttir, Nigel Gilbert, and Klaus Moessner. 2017. Recognising Activities at Home: Digital and Human Sensors. In *Proceedings of ICFNDS '17, Cambridge, United Kingdom, July 19-20, 2017,* 11 pages.
DOI: 10.1145/3102304.3102321

ICFNDS '17, Cambridge, United Kingdom

DOI: 10.1145/3102304.3102321

1 INTRODUCTION

Social researchers have a great interest in household practices, among other things, family dynamics and child-rearing (e.g. [31]; [13]), practices around meals [24], sleep [54], assisted living arrangements and mobile health solutions (e.g. [35]; [36]), homeworking [49] and energy-related practices [40]. Existing social research methods are both qualitative and quantitative, and often some combination of the two are used for pragmatic and constructivist purposes [45].

Qualitative methods are used to acquire rich in-depth data. Observations and open-ended interviews are particularly effective in capturing the meanings participants attach to various aspects of their everyday lives and relations (e.g. [23]). Quantitative methods such as questionnaires and surveys capture qualitative information in formalised ways for computational processing, and are widely used in large scale studies on demographics, household economics and social attitudes (e.g. [1], [4]). Time-use diaries are also used to log activity sequences [17], and to seek evidence of life changes and social evolution [25]. Efforts to harmonise time use surveys across Europe have delivered guidelines (HETUS) on activity coding for analysing the time use data [19], but interviews and observations are commonly used to cross-validate what goes on, and to calibrate and amplify the meaning of the diary evidence, including the use of activity sensors and video cameras [28].

With the advance of sensor technologies, researchers are provided with new ways of capturing activities at home. For example, temperature, humidity and light sensors provide information about the physical environment where activities occur, energy monitors report information about the electrical devices used to assist the activities, and accelerometers capture the motion of the people who are performing these activities. Such rich contextual information has attracted social researchers. For example, Williams et al. [54] discuss the use of accelerometers to study people's sleep patterns. Amft and Tröster [12] study people's dietary behaviour by using inertial sensors to recognise movements, a sensor collar for recognizing swallowing and an ear microphone for recognizing chewing. Wang et al. [52] help in detecting elderly accidental falls by employing accelerometers and cardiotachometers.

Sensor-generated data is becoming widely available and the topic of activity recognition [14] has thrived in recent years with applications in areas such as smart homes and assisted living. Researchers have investigated activity recognition methods using data obtained from various types of sensors, for instance, video cameras [41], wearables [32] and sensors embedded in smartphones [46]. Numerous activity recognition algorithms are proposed in the literature, mainly based on the assumption that by sensing the environment it is possible to infer which activities people are performing. Bayesian Networks (BNs) and Hidden Markov Models (HMMs) [26] are popular methods due to their ability to recognise latent random variables in observing sequences of sensor-generated data. Other approaches rely on Conditional Random Fields (CRFs) [50] and Artificial Neural Networks [44]. A more detailed discussion will be given in section 2.

To evaluate the adequacy of inferences about activities derived from sensor data, it is necessary to have a record of what activities are occurring from direct observation to obtain the so-called 'ground truth'. In the literature, there are three main types of approach. The first relies on video cameras to record what participants are doing during an experiment. For example, Lin and Fu [34], use multiple cameras and floor sensors to track their participants. Although the data quality can be guaranteed in a controlled lab, this method is very intrusive and difficult to deploy in an actual home. A second common way of establishing ground truth is by asking participants to carry out a predefined list of activities, again, in a controlled environment. For example, Cook et al. [20] ask their participants to carry out scripted activities, predetermined and repeatedly performed. Both of these methods correspond with social research methods, such as questionnaires, surveys and interviews, in generating what Silverman calls 'researcher-provoked' data [47]. The outcomes may suffer the bias introduced by the researchers in provoking participants' activities as opposed to observing them without interference. The third type of approach relies on human annotators to label sensor-generated data manually. For example, Wang et al. [53] first conducted a survey with their participants to have a self-reported record of their main activities, then an annotator both logged and annotated the activities performed by the participants in a living lab environment where video cameras were also used to record what the participants were doing. This type of approach relies heavily on the annotator's knowledge of participants' activities and their understanding of participants' everyday practices, but may also be challenged by discrepancies between the research-provoked survey and video data and non-provoked sensor-generated data.

Audiovisual recorders and digital sensors do not generate researcher-provoked data as do interviews, questionnaires, surveys and controlled experiments. Arguably, the same can be said about time use diaries. Following the HETUS guidelines, participants use their own words to describe primary and secondary activities and they are prompted simply in that one activity follows another sequentially. This neutral way of requesting the record gives no idea of researchers' interests, making it simpler to give truthful accounts than it is when answering stylised questions. However, the diary format can be stylised and simplified when the interface is *appified* [48]. This approach might be useful in securing data on researcher-defined activity categories, although risking significant loss of participants' own definitions of what they do.

In this study, we followed the HETUS guidelines in using time-use diaries as a test for establishing ground truth against the automated recognition of activity types. Activity recognition methods are applied to data generated by 'digital sensors', while data on activities are captured simultaneously in time-use diaries provided by 'human sensors'. To what extent do digital sensors agree with human sensors, i.e., how can we quantify the agreement between what we learn from sensor-generated data and what we know from self-reported unprovoked data about activities in the home?

To contribute to answering this question, we conducted a trial in a residential house, collecting data from a set of sensors and from a time use diary recorded by one of the two occupants over four consecutive days. The sensors captured temperature, humidity, range (detecting movements in the house), noise (decibel levels), brightness of light and energy consumption. In using the sensor-generated data for activity recognition, we adopt an unsupervised learning approach based on a Hidden Markov Model, and investigate the application of mean shift clustering [16] and change points detection [39] for constructing features. Furthermore, we propose a method for measuring the agreement between activity sequences proposed by the activity recognition algorithm and those reported by the participant, based on the Levenshtein distance [33].

The contributions of this paper are two-fold. First, we present a new data collection framework for recognising activities at home, i.e., a mixed-methods approach of combining computational and qualitative types of non-provoked data: sensor-generated and time use diary. Secondly, we propose an evaluation method for measuring the agreement between the sensor-supported activity recognition algorithms and the human constructed diary.

The rest of the paper is organised as follows. In section 2, we discuss related work. In section 3, we give an introduction to the home setting. Thereafter, in section 4, we describe the data collected for this study, including both the sensor data and the time use diary data. In section 5, we show the features we construct and introduce our activity recognition algorithm. In section 6, we present the metric for evaluating agreement between activities recognised by the sensor-generated data and what is reported by the participant. Finally, we provide a discussion in section 7.

2 RELATED WORK

In this section, we discuss the works that have been recently published in the area of automated activity recognition in home-like environments in terms of the sensors they use, the activities they detect, and the recognition methods they adopt or propose.

Lin and Fu [34] developed a three-layer multi-user preference model for service provision. At the first layer, K-Means clustering and domain knowledge are used to create context out of raw data. The second layer learns K Dynamic Bayesian Networks (DBNs), each of which model electrical appliance (EA) controllers for residents' preferences. The third layer uses a Bayesian Network to meet the needs of each resident, thus learning a "group service" corresponding to the configuration of the K EA controllers. The authors show that in their smart home laboratory setting consisting of location tracking using a camera and floor sensors, infrared motion detectors, temperature and light sensors, the recognition accuracy they achieve is 89% for 5 activity types: turning on TV, turning on the lights in both living room and study, playing music and turning on a table lamp in a multi-resident scenario with three students where the activity of each user will influence the activities of the others.

Cook et al. [20] reported on the detection of social interactions in smart building environments. Their dataset was collected during the CASAS smart environment project [43] from a three bedroom apartment and workplace testbeds. Their experiment reveals 8 activity types between residents (play checkers, fill a medication dispenser, hang up clothes, move the couch, water plants, sweep kitchen floor, prepare dinner, pay bills). The data originates from motion and temperature sensors, water and stove (ad-hoc) usage sensors, energy monitors and lighting controls and contact detectors for cooking pots, phone books and medicine containers. The authors use HMMs for identifying activities and interactions between people in a multi-person setting.

Hsu et al. [27] investigated the problem of activity recognition in a multiple-resident environment based on Conditional Random Fields (CRFs). They evaluated their approach with several strategies (iterative inference and decomposition inference) against the CASAS dataset for a set of 15 individual and cooperative activities, and found that data association of non-obstructive sensor data is important to improve the performance of activity recognition in a multiple-resident environment. In addition, they show that the models achieve a higher accuracy with raw data rather than processed data.

Chiang et al. [18] proposed three different models based on HMMs and DBNs. A Parallel HMM is designed based on an independent HMM for each resident. A Coupled HMM is then proposed which assumes that the activities of different residents are dependent. Finally, A DBN consisting of the Coupled HMM plus additional interaction features and vertices to model cooperation is created. An evaluation against the CASAS dataset reports a higher accuracy for the DBN based on a raw feature set than that achieved by Hsu et al. [27].

Wang et al. [53] evaluated Coupled HMMs and Factorial CRFs based algorithms on a multi-modal wearable sensor platform in a smart home setting, to capture location, user movements (with accelerometers), human-object and human-human interaction with voice recognition, and environmental information on temperature, humidity and light. During their experiment, they also collected survey reports on where, when, who and for how long activities were performed by 30 students as well as the experience of their interactions. They report high accuracy in recognising a set of 21 activities and an improvement by applying a Correlation-based Feature Selection algorithm. The improvement was shown to be more significant for single-user activity. The set of 21 activities includes 14 individual activities such as brushing teeth and hair, washing face, making pasta, coffee or tea, toileting, ironing, vacuuming, using a phone, computer or TV, reading, having a meal and drinking. Seven activities were cooperative: making pasta or coffee, cleaning the table, queueing for the toilet, watching TV and using a computer.

Fang et al. [22] also employed the CASAS dataset to train a neural network with back propagation in order to recognise activities. In addition, the authors implement an inter-class distance feature selection algorithm capable of identifying the best K feature subset to be presented at the input layer of the neural network. Their extensive evaluation against Naive Bayes classifier-based and HMM-based solutions shows that their neural network has a better recognition accuracy than these. The set of activities they detect includes bed-to-toilet, having breakfast, lunch or dinner, sleeping, working, doing the laundry, leaving home, taking medicines and night wandering.

Prossegger and Bouchachia [42] instead used the ARAS dataset [11] in order to learn a classification of activities of multiple residents in two houses with a month of labelled data. Their work proposes an extension of the incremental decision tree ID5R, named E-ID5R, where three phases consisting of tree construction, classification of new instances and tree evolution are modelled. Their dataset contains 27 possible activities (see [11] for the entire list) described by the measurement of force sensitive, contact, proximity, temperature and sonar distance sensors, infrared receivers, photocells and pressure mats.

The work by Ordonez et al. [37] evaluated transfer learning with HMMs that uses prior accumulated experience on new target houses where little data is annotated. The dataset is generated with motion-sensitive passive infrared sensors, reed switches for doors and cupboards and float sensors to measure the toilet being flushed, with a target set of 7 activities: leaving, toileting, showering, sleeping, having breakfast, eating dinner and drinking. In order to evaluate transfer learning, the authors propose a meta-features representation for each sensor depending on its location, leading to a slightly lower accuracy in a single house scenario but higher accuracy in transfer scenarios with little annotation.

Chen and Tong [15] proposed a two-stage method based on HMM and CRF in order to improve accuracy of activity recognition using a "combined label" state definition. Since the authors advocate the need to recognise parallel activities of residents either performed independently or collaboratively, they define the labels in order to fit such needs. While evaluating using the CASAS dataset with the usual set of 15 activities and up to 27 different bi-dimensional labels, they obtain an improvement in accuracy with respect to previous models. Moreover, they obtain a very high recognition accuracy when multi-label classification is considered.

Fan et al. [21] provided a comparative study of four machine learning algorithms for activity recognition in home-like environment. The sensors they used include grid-eye infrared array sensors, force sensors, noise sensors, and electrical current detectors. For their experiment, the participants were asked to perform a predefined list of activities in a home lab, including eating, watching TV, reading books, sleeping and visiting friends. Based on an evaluation of the four algorithms for the activity recognition task, the authors found that a straightforward meta-layer network model outperforms other models.

van Kasteren et al. [51] introduced a sensor and annotation system for activity recognition in a home setting. The sensors they used are 14 state-change sensors that were placed on the doors, cupboards, refrigerator and toilet flush. The activities that were recognised include leaving, toileting, showering, sleeping, having breakfast, eating dinner, and drinking, annotated by the participants themselves. Two probabilistic models, HMM and CRF, were investigated for the activity recognition task.

Kelly et al. [28] tested the feasibility of using wearable cameras to validate time use diaries. Participants were asked to wear a camera and at the same time keep a record of time use over a 24-hour period. During an interview with each participant afterwards, the visual images were used as prompts to reconstruct the activity sequences and improve upon the activity record. No significant differences were found between the diary and camera data with respect to the aggregate totals of daily time use. However, for discrete activities, the diaries recorded a mean of 19.2 activities per day, while the image-prompted interviews revealed 41.1 activities per day.

While several other approaches exist for activity recognition and capture, they mostly employ only wearable sensors (i.e. see Lara and Labrador for a recent survey [32]), and thus cannot be applied in multi-modal scenarios of smart-home settings with fixed, unobtrusive and privacy preserving ambient sensors. In addition, due to the time-series nature of activity recognition in the home environment, supervised algorithms not incorporating the notion of temporal dependence might lead to poor performance in activity recognition, so such work is not reviewed here.

The types of sensors used in the aforementioned research range from fixed sensors such as motion detectors and cameras to wearables such as accelerometers that differ in measurement, intrusiveness and price. The types of activities that can be recognised are also largely determined by the types of sensors used. Probabilistic models (e.g., HMM, CRF) are widely used for activity recognition. As for establishing ground truth, three methods are most common: using video cameras, manual annotation and a predefined list of activities.

In this work, we use a suite of fixed and unobtrusive sensors. For activity recognition, we build our model based on HMMs. In particular, we investigate the use of mean shift clustering and change points detection techniques for feature construction. Our work differs from similar studies in that we adopt a mixed-methods approach for the problem of recognising activities at home, and we evaluate its effectiveness using a formal framework.

3 EXPERIMENT SETTING

For this work, we installed a suite of sensors in an apartment. The data collected by the sensors was encrypted and sent to a central server over the internet.

3.1 Sensor Modules

We used six types of sensor modules, as summarised in Table 1.

The first five sensor modules are encapsulated in a sensor box, as shown in Figure 1 (a), coordinated by a Seeeduino Arch-Pro [6]. The temperature and humidity sensor HTU21D [7] is managed via an I2C interface and sampled periodically by the client application deployed on the ARM core. An Avago ADPS-9960 light sensor [9] , also managed via an I2C interface, is used to sample ambient light measured in $\frac{\mu W}{cm^2}$. The GP2Y0A60SZ ranging sensor from Sharp [5] is an analog sensor with a wide detection range of 10 cm to 150 cm

Table 1: Sensor Modules

Sensor modules		Measurement
Sensor Box	Temperature sensor	℃
	Humidity sensor	%
	Light Sensor	$\frac{\mu W}{cm^2}$
	Ranging sensor	*cm*
	Microphone	*dB SPL*
Energy monitor		*watts*

and an update rate of 60 Hz, which is sampled via a 12 bit ADC and converted through the manufacturer's calibration table. Finally, the MEMS Microphone breakout board INMP401 [8] is used to sample noise levels in the environment via an ADC and the values are converted to decibels (dB SPL).

The other sensor module used in this work is a commercial electricity monitoring kit from CurrentCost [2], as shown in Figure 1 (b). It features a CT clamp, a number of individual appliance monitors (IAMs) and a transmitter to measure the energy consumption in watts of the whole house as well as the individual appliances.

(a) Sensor Box **(b) Electricity Monitor**

Figure 1: Sensor Modules

Compared to the works discussed in Section 2, the sensor modules used for this work are less obtrusive and require no effort on part of the participants.

3.2 Trial Home

The trial home is an apartment consisting of a living room, a bedroom, a kitchen, and a bathroom, and is occupied by a couple. With the formal consent of the participants, we distributed a number of sensors in the rooms of the home.

The sensor distribution used for this work is as follows. The CT clamp is connected to the electricity main to measure total electricity consumption. In the Living room we placed a sensor box near the entrance of the room, to capture motion in and out along with the environmental variables. Two IAMs were plugged in to capture the use of devices in routine use such as plugged-in laptops, a router and cellphone chargers. In the Bedroom we placed a sensor box at the entrance of the room. In the Kitchen we placed a sensor box near the cooking hob with the aim of capturing cooking activities. As gas is the main energy source for cooking, IAM could not be used to monitor energy consumption of cooking devices. No sensors were installed in the bathroom, as it was considered too intrusive.

4 DATA SETS

Two types of data were collected, sensor-generated data and a time use diary. The data sets presented here cover a period of 4 consecutive days in December 2016 from 6:00am on the first day until 11:50pm on the last.

4.1 Sensor-generated Data

The sensor-generated data consists of six types of readings in accordance with the six sensor modules shown in Table 1.

The data was collected from the sensor boxes every 3 to 5 seconds. A data sample from a sensor box is:

{'Box_ID': 123, 'Timestamp': 2016-12-13 09:00:00,'Temperature': 20, 'Humidity': 50, 'Sound': 45, 'Range': 100, 'Light': 583}

For *range*, the sampling rate is 100Hz and the lowest value , i.e. the distance to the nearest detected moving object, is used as the datum in each collection cycle so as to minimise false negatives in detecting movements. Noise level is sampled by an on-device conversion of air pressure changes at around 3000Hz.

Data is captured by the electricity monitors (IAMs) every 5 minutes. A data sample from an IAM is:

{'IAM_ID': 123, 'Timestamp': 2016-12-13 09:00:00,'Watts': 100}

In total, there are $(104145 + 104156 + 104159) \times 5$ data samples from the three sensor boxes with respect to the five sensor modules and 1075×3 data samples from the CT clamp and two IAMs.

4.2 Time Use diary

During the four days of the experiment, one of the two occupants was asked to keep a diary of time use based on the HETUS model [19], recording activities with an interval of 10 minutes. Table 2 gives an idea of what is recorded. In total, we have 540 data points from the diary over the course of 4 days.

Table 2: Time use diary example

Time	Activity	Location	Devices
17:50-18:00	Entertaining	Living room	Laptop
18:00-18:10	Cooking	Kitchen	Oven
18:10-18:20	Cooking	Kitchen	Oven
18:20-18:30	Dining	Living room	Laptop

A time-use diary may misreport in several ways. First, the start and end time of individual activities may not be accurately recorded, i.e., either earlier or later than the actual occurrence. This is called *time shifting*. Secondly, there might be activities that occurred but were not recorded, which are *missing values*. Thirdly, since the time use diary recorded only one person's activities, it only provides partial information about what was happening in the house.

We focus here on the four types of activities that are most frequently performed according to the diary: cooking, dining, entertaining and sleeping. Table 3 gives a summary of how many times each occurred and how much time was spent on them.

Table 3: Number of occurrences and time spent for each type of activities in the data set

Activities	Number of occurrences	Percentage of time
Cooking	12	10.74%
Dining	6	3.89%
Entertaining	13	31.85%
Sleeping	5	33.33%

5 RECOGNISING ACTIVITIES

Sensor-generated data provides a digital means of looking into the life of a household. Such a window in itself does not tell directly what activities take place but it provides rich contextual information drawn from the aggregate of environmental variables. Our objective in this section is to investigate what kinds of features can be drawn from the sensor-generated data and how such features can be used for activity recognition.

5.1 Feature Construction

Activities give rise to change in sensor readings. For example, when cooking, the *temperature* may rise in the kitchen because of the heat emitted from the hob, *humidity* levels go up and *range* readings may fluctuate quite intensively because of the physical movements involved. These types of changes in the sensor-generated data are essential to better understand the context of activities and to recognise their occurrence.

There are two types of patterns in the sensor readings that are useful in identifying activities. The first type is clustering, i.e., absolute values of sensor readings appear naturally in clusters. For example, the readings of the ranging sensor are either the maximum value during periods when nothing comes in and out of range or distinctly much smaller values. The second type relates to the distribution of sensor readings along the time line, thus taking into account both time dependency and value differences between sensor readings.

Accordingly, we investigate the application of two methods for constructing features from sensor-generated data. The first, *mean shift*, aims at clustering the readings of sensor data into different value bands. The second method, *change points detection*, aims at finding meaningful points of change in the sequences of sensor-generated data.

5.1.1 Preprocessing. To align with the time use diary, we re-sample the sensor data with bins of 10 minutes so that each data point has a corresponding activity label. Re-sampling is done using the maximum values for temperature, humidity, brightness, noise level, and the minimum values for range. This resampling yields 540×3 data points for each type of sensor reading.

5.1.2 Mean shift. Mean shift is a non-parametric clustering method that does not require prior knowledge of the number of clusters. It is based on an iterative procedure that shifts each data point to its nearest local mode, by updating candidates for centroids to be the mean of the data points within its neighbourhood [16].

Given a set of data points S in a n-dimensional Euclidean space X, mean shift considers these data points as sampled from some underlying probability density function. In this work, we chose to use a flat kernel K with a bandwidth h for the estimation of the probability density function, as defined below:

$$K(x) = \begin{cases} 1 & \text{if } \|x\| \leq h, \\ 0 & \text{otherwise.} \end{cases}$$

The sample mean at $x \in X$ is

$$m(x) = \frac{\sum_{s \in S} K(s-x)s}{\sum_{s \in S} K(s-x)}$$

The difference $m(x) - x$ is called mean shift and the mean shift algorithm is the procedure of repeatedly moving data points to the sample means until the means converge. In each iteration, s is updated by $m(s)$ for all $s \in S$ simultaneously. The implementation is based on python scikit-learn [38].

As an example, Figures 2, 3 and 4 compare the results of processing with the mean shift clustering algorithm for the range and noise-level readings in the kitchen and the electricity readings in the living room. The readings for both range and noise in the kitchen generate two clusters. For range, a straightforward explanation is that one cluster represents the times when no movements are detected in the kitchen and the other cluster represents the times when movements are detected. As for the noise, the two clusters represent the times when a lot of noise is detected in the kitchen and when the kitchen is relatively quiet. For the electricity readings, four clusters are generated that represent different levels of energy consumption.

Figure 2: Mean shift clustering of range readings from a sensor box in the kitchen

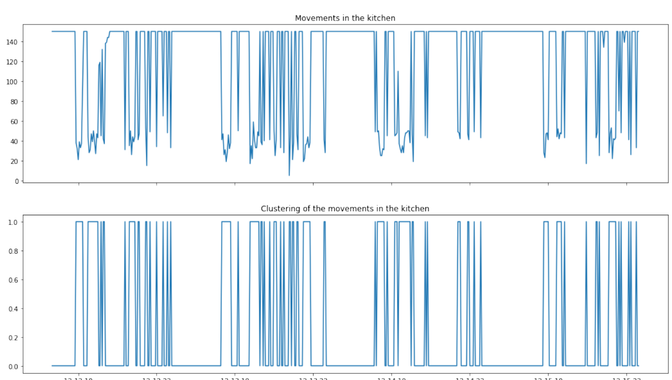


Figure 3: Mean shift clustering of noise-level readings from a sensor box in the kitchen

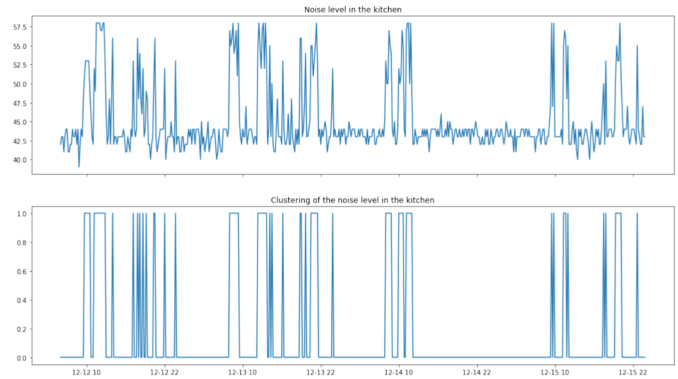
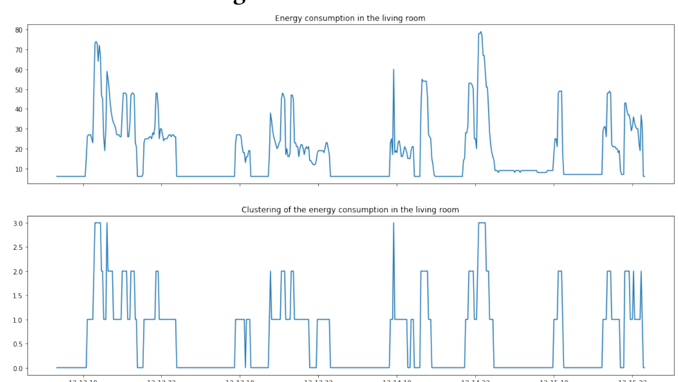


Figure 4: Mean shift clustering of electricity readings from the IAMs in the living room



5.1.3 Change points detection. Change points detection is a method of estimating the times at which the statistical properties of a sequence of observations change [39].

Given a sequence of data, $x_{1:n} = (x_1, \ldots, x_n)$, a change is considered to occur when there exists a time $\tau \in \{1, \ldots, n-1\}$ such that the statistical properties of $\{x_1, \ldots, x_\tau\}$ differ from that of $\{x_{\tau+1}, \ldots, x_n\}$, e.g., in *mean* or *variance*. In the case of multiple changes, a number of change points $\tau_i, i \in \{1, \ldots, m\}$ are identified, which split the sequence of data into $m+1$ segments.

A lkelihood based framework is widely used for change points detection. The most common approach for detecting multiple change points is to minimise

$$\sum_{i=1}^{m+1} [C(x_{(\tau_{i-1}+1):\tau_i})] + \beta f(m)$$

where C is a cost function for assuming a change point at τ_i in the time series data and $\beta f(m)$ is a penalty function to avoid over fitting (i.e., too many change points).

For our sensor data, we focus on detecting the changing *mean* in the sensor readings. The cost function is the negative log-likelihood. A manual setting is used for the penalty function so that the number of change points can be adjusted. The change points detection algorithm is the pruned exact linear time (PELT) [30] which is computationally efficient and provides an exact segmentation. This is implemented using the R package [29]

As an example, Figures 5 and 6 show the results of change points detection for temperature and humidity in the kitchen.

Figure 5: Change points of temperature readings from a sensor box in the kitchen

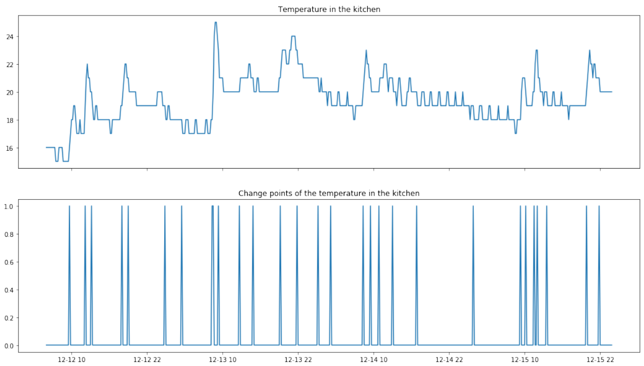
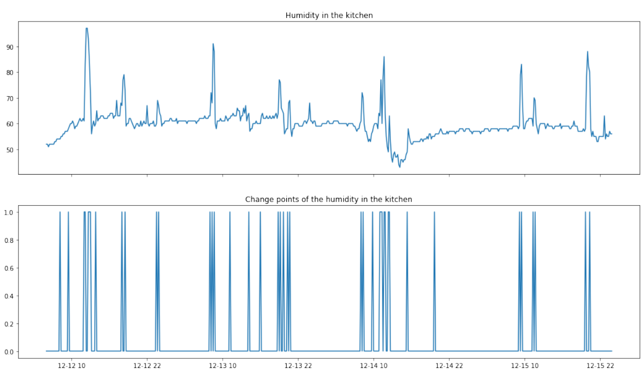


Figure 6: Change points of humidity readings from a sensor box in the kitchen



The gap between change points can be used to identify activities. For instance, when the house is asleep we can expect a long gap between changes in electricity consumption and brightness, from the time of going to bed to getting up again, as shown in Figures 7 and 8.

Figure 7 shows big gaps in total energy consumption, indicated by the width and height of the bars, detected between midnight and early morning every night. A similar and aligned pattern can be found in the brightness of the bedroom.

5.2 Recognition Method

Hidden Markov models (HMMs) have proven to be effective in modelling time series data [55]. They are a good fit for recognising activities from sensor data in the sense that they are capable of recovering a series of states from a series of observations.

An HMM is a Markov model whose states are not directly observable but can be characterised by a probability distribution over observable variables. In our case, the hidden states correspond to the activities performed by the participant and the observations correspond to the sensor readings. There are two assumptions in HMMs, as illustrated in Figure 9. The first is that the hidden state y_t

Figure 7: Gaps between change points in the electricity readings from the CT clamp connected to the electricity main

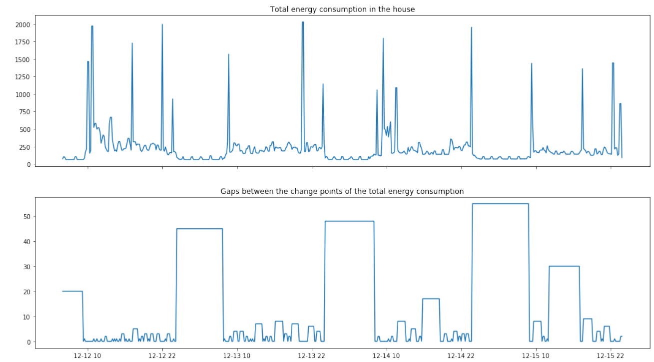
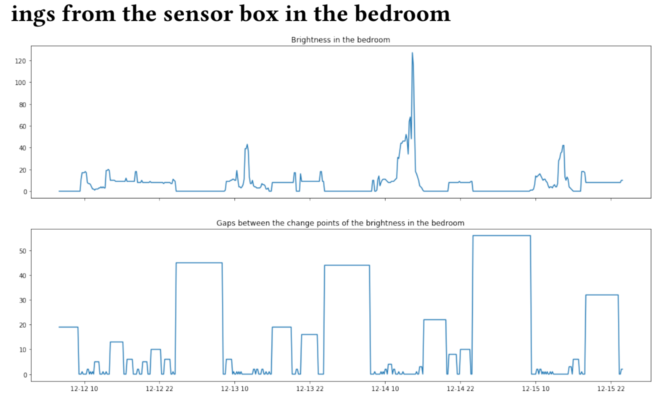
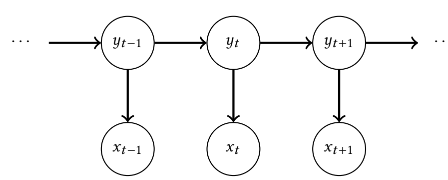


Figure 8: Gaps between change points in the brightness readings from the sensor box in the bedroom



(at time t) depends on the previous hidden state y_{t-1}. The second is that the observation x_t depends on the hidden state y_t.

Figure 9: Graphical representation of a Hidden Markov Model



An HMM is specified using three probability distributions: (i) the initial state probability distribution, (ii) the transition probability of moving from one hidden state to another, and (iii) the emission probability of a hidden state generating an observation. The parameters of these three probability distributions can be estimated by maximising the joint probability:

$$P(y, x) = P(y_1)P(x_1|y_1)\prod_{t=2}^{T} P(y_t|y_{t-1})P(x_t|y_t)$$

We focus on four types of activities, **cooking**, **dining**, **entertaining** and **sleeping**, as shown in Table 3. For each, we built HMMs using combinations of features constructed by the methods presented in Section 5.1. For each HMM, there are two hidden states: either a particular activity is occurring or not occurring. Since the features are of discrete values, HMMs with multinomial emissions are used. In the next section, we describe how sequences of hidden states returned by the HMMs are related to the sequences of activities recorded in the time use diary. The implementation is based on python hmmlearn [3].

6 AGREEMENT EVALUATION

6.1 Evaluation Metric

In the previous section, we introduced the activity recognition framework. By feeding the sensor data into the HMMs, sequences of hidden states can be extracted. The problem then is how we can evaluate the agreement between sequences of discrete state generated by the HMMs and the activity sequences recorded in the time use diary.

As we discuss in Section 4.2, the activity sequences recorded in the diary may contain time shifts and missing values. Direct comparison may exaggerate the dis-similarity introduced by such noise. Thus, we need an agreement evaluation metric that is able to alleviate the effect.

Another issue for the agreement evaluation is that the labels of the hidden states are not directly mapped to the activity labels, i.e., the hidden states cannot be prescribed for any particular activity. Therefore, for each type of activity, we evaluate all the possible mappings between the hidden states and the activity labels.

A suitable metric for this task is the Levenshtein distance (LD) [33] which has been widely used for measuring the similarity between two sequences. It is defined as the minimum number of insertion, deletion or substitution operations needed to transform one sequence into the other.

Formally, given two sequences s and q, the Levenshtein distance between these two sequences $D_{s,q}(|s|, |q|)$ is defined by

$$D_{s,q}(i, j) = \begin{cases} max(i, j) & \text{if } min(i, j) = 0, \\ min \begin{cases} D_{s,q}(i-1, j) + 1 \\ D_{s,q}(i, j-1) + 1 \\ D_{s,q}(i-1, j-1) + 1_{(s_i \neq q_j)} \end{cases} & \text{otherwise.} \end{cases}$$

where $i \leq |s|, j \leq |q|$; $1_{(s_i \neq q_j)}$ is an indicator function that equals to 0 when $s_i = q_j$ and equals to 1 otherwise. The three lines in the *min* bracket respectively correspond to the three operations transforming s into q, i.e., deletion, insertion, and substitution (depending on whether the respective elements are the same). The costs of the three types of operations in the standard Levenshtein distance are all set to be 1.

The inputs to the Levenshtein distance, in our case, are respectively a sequence of activity labels generated by the HMMs and a sequence of activity labels recorded in the time use diary. We also attempt to alleviate the effect introduced by the time shifting and missing values when evaluating the (dis-) similarity between the two sequences. More specifically, the transformation cost should be considered less when the dis-similarity between the two sequences is mainly caused by time shifting.

For this purpose, we make an adjustment to the cost associated with the three types of operations in the Levenshtein distance: the costs of inserting and deleting 0 (indicating a particular activity is not performed) are set to 0.5, and the costs of inserting and deleting 1 (indicating a particular activity is performed) are set respectively at 0.8 and 1.0. The value difference between these costs is mainly used to differentiate the penalty of different operations, while the influence of the exact value difference will be investigated in future work. In this way, the output from the Levenshtein distance is the minimum cost of the operations that are needed to transform one sequence to the other. The implementation is based on the python package weighted-levenshtein [10].

6.2 Analysis

For each of the four types of activity, we fitted the HMMs a thousand times with different combinations of features using randomised initial states. Table 4 lists the set of features that achieves the best agreement in terms of the adapted Levenshtein distance (LD) between the activity sequences generated by the HMMs and that recorded in the time use diary. The prefixes, *MS_*, *CP_* and *Gap_CP_*, represent the mean shift clustering results, the change points detection results, and the gaps between the detected change points of a particular type of sensor reading. The features are associated only with the sensors placed in the room where the corresponding activities occur, as specified by the sensor location in table 4.

Table 4: Optimal Recognition Features for Types of Activities

Activities	**Sensor Location**	**Features**	**LD**
Cooking	Kitchen	CP_Temperature, CP_Humidity, MS_Range, MS_Sound	38.4
Dining	Living room	CP_Sound, Gap_CP_Electricity	36.0
Entertaining	Living room	MS_Electricity, MS_Sound, MS_Light, MS_Range	61.1
Sleeping	Bedroom, Electricity main	Gap_CP_Electricity, MS_Sound, MS_Light, MS_Range	12.6

Table 4 shows that, for recognising cooking activities, the best agreement is achieved using the features, *CP_Temperature*, *CP_Humidity*, *MS_Range* and *MS_Sound*. This is confirmed in an interview with the participant, who talked about often cooking hot meals three times a day, using the oven, cooking pans and a kettle. The participant also mentioned listening to loud music while cooking. For dining activities, the best agreement is achieved using the features of *CP_Sound* and *Gap_CP_Electricity*. From the interview, we know that the participant always watches videos on a plugged-in laptop while dining, which may lead to changes in noise levels as well as in the electricity usage in the living room. For entertaining activities, the best agreement is achieved using the combination of features, *MS_Electricity*, *MS_Sound*, *MS_Light*, and *MS_Range*. From the interview, we know that the participant uses a laptop for entertainment at home and thus the electricity usage measured by the corresponding IAM will stay steady at a higher level. For sleeping, the best agreement is achieved using the features, *Gap_CP_Electricity*, *MS_Sound*, *MS_Light*, and *MS_Range*. This corresponds to the fact that most devices are turned off during sleep. Lights are off, noise levels keep steady at a low level of background environmental noise, and there is little if any motion detected.

For illustration, we plot four maps, one for each type of activity sequence (figures 10, 11, 12 and 13). In each figure, the upper part shows the state sequences generated by an HMM (using the specific set of features) and the lower part shows the activity sequences recorded in the time use diary. The black bins represents the time slices when a particular activity is performed.

Figure 10: Cooking

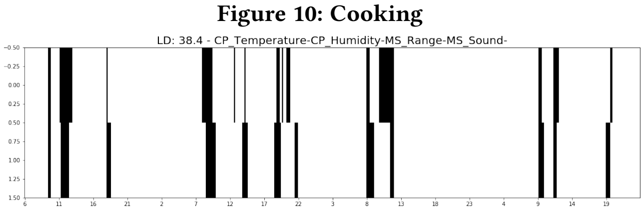


Figure 11: Dining

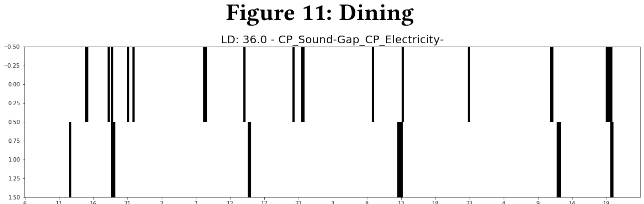


Figure 12: Entertaining

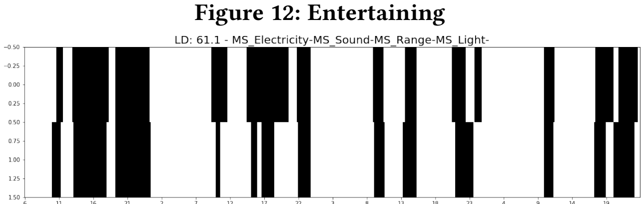
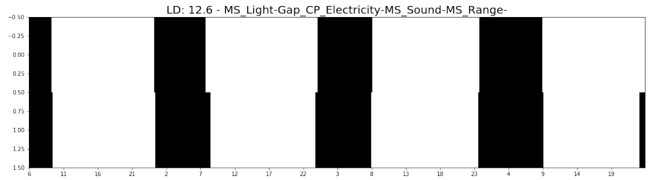


Figure 13: Sleeping



The activity durations generated by the HMMs overlap with those recorded in the time use diary, with some local shifts along the time line. The only exception is dining, which suggests that the features constructed are not sufficient to distinguish dining from simply entertainment. Dining activities are sparse. They do not have unique enough indicators, one reason being that dining occurs with other activities such as entertainment.

7 DISCUSSION

In this paper, we have presented a mixed-methods approach for recognising activities at home. In particular, we investigated ways of extracting features from sensor-generated data for activity recognition. We also proposed a method for evaluating the agreement between the predicted activities from models trained by the sensor data and the activities recorded in a time use diary.

The focus of this work is not on improving the recognition performance of particular models but to present a framework of quantifying how activity recognition models trained by sensor-generated data can be evaluated on the basis of their agreement with the activities recorded in time use diaries. Such a framework may be useful in several ways. First, the evaluation results can provide evidence about which types of sensors are better for detecting certain types of activities in a household. This may further help in understanding how certain activities affect the environment. For example, if change points in noise-levels point to better agreement for dining activities, it is very likely that a household is dining in a loud environment. Secondly, the agreement between the sensor-generated data and the time use diary can reflect the quality of the diary, especially when further compared with information obtained in interviews with participants about their life at home. Thirdly, the evaluation framework can also be generalised for experimental settings that use other means than records of time use to obtain information on household activities, e.g., questionnaires or surveys.

This is an on-going research that is investigating the use of digital sensors for social research, using household practices as a testbed. As this is written, we are in the process of recruiting and collecting data from three types of households: single occupant, families with children and 2+ adults. There are several directions to consider for future work. We are adding a wearable wristband sensor to the setting to detect the proximity of participants to each sensor box via bluetooth RSSI (received signal strength indicator). Such data will give us a more accurate reading of presence and co-presence of particular occupants in different parts of their home, while also helping us in obtaining more accurate start and end times of certain activities. We will continue to investigate other activity recognition methods and feature selection techniques. Also, we are interested in employing post and assisted labelling mechanisms, for example, by asking participants to assign an agreement score to the activity sequences generated by our activity recognition models. In this way, another layer of agreement can be added to the evaluation.

ACKNOWLEDGMENTS

The authors thank Dr. William Headley for the design and manufacture of the sensor box (also known as desk egg).

The work was carried out as part of the 'HomeSense: digital sensors for social research' project funded by the Economic and Social Research Council (grant ES/N011589/1) through the National Centre for Research Methods.

REFERENCES

[1] Online. British Social Attitudes Survey (33rd edition). Accessed : 2017-30-3. (Online). http://www.bsa.natcen.ac.uk/latest-report/british-social-attitudes-33/introduction.aspx.
[2] Online. Current Cost - Individual Appliance Monitor (IAM). Accessed : 2017-30-3. (Online). http://www.currentcost.com/product-iams.html.
[3] Online. hmmlearn. (Online). http://hmmlearn.readthedocs.io/.
[4] Online. Office for National Statistic (UK). Accessed : 2017-30-3. (Online). https://www.ons.gov.uk/.
[5] Online. Pololu Carrier with Sharp GP2Y0A60SZLF Analog Distance Sensor. Accessed : 2017-30-3. (Online). https://www.pololu.com/product/2474.
[6] Online. Seeeduino Arch-Pro. Accessed : 2017-30-3. (Online). https://developer.mbed.org/platforms/Seeeduino-Arch-Pro.
[7] Online. SparkFun Humidity and Temperature Sensor Breakout - HTU21D. Accessed : 2017-30-3. (Online). https://www.sparkfun.com/products/retired/12064.
[8] Online. SparkFun MEMS Microphone Breakout - INMP401 (ADMP401). Accessed : 2017-30-3. (Online). https://www.sparkfun.com/products/9868.
[9] Online. SparkFun RGB and Gesture Sensor - APDS-9960. Accessed : 2017-30-3. (Online). https://www.sparkfun.com/products/12787.
[10] Online. weighted-levenshtein. (Online). http://weighted-levenshtein.readthedocs.io/.
[11] H. Alemdar, H. Ertan, O. D. Incel, and C. Ersoy. 2013. ARAS human activity datasets in multiple homes with multiple residents. In *2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops*. 232–235.
[12] Oliver Amft and Gerhard Tröster. 2008. Recognition of dietary activity events using on-body sensors. *Artificial Intelligence in Medicine* 42, 2 (2008), 121 – 136.
[13] R. Atkinson and K. Jacobs. 2016. *House, Home and Society*. Palgrave.
[14] Asma Benmansour, Abdelhamid Bouchachia, and Mohammed Feham. 2015. Multioccupant Activity Recognition in Pervasive Smart Home Environments. *ACM Comput. Surv.* 48, 3 (2015), 34:1–34:36.
[15] Rong Chen and Yu Tong. 2014. A Two-stage Method for Solving Multi-resident Activity Recognition in Smart Environments. *Entropy* 16, 4 (2014), 2184–2203.
[16] Yizong Cheng. 1995. Mean Shift, Mode Seeking, and Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17, 8 (1995), 790–799.
[17] Alain Chenu and Laurent Lesnard. 2006. Time Use Surveys: a Review of their Aims, Methods, and Results. *European Journal of Sociology* 47, 3 (2006), 335–359.
[18] Yi-Ting Chiang, K. C. Hsu, C. H. Lu, Li-Chen Fu, and Jane Yung-Jen Hsu. 2010. Interaction models for multiple-resident activity recognition in a smart home. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 3753–3758.
[19] European Communities. 2009. *Harmonised European time use surveys: 2008 guidelines*. Eurostat Methodologies and Working Papers. Population and social conditions.
[20] Diane J. Cook, Aaron Crandall, Geetika Singla, and Brian Thomas. 2010. Detection of Social Interaction in Smart Spaces. *Cybernetics and systems* 41, 2 (2010), 90–104.
[21] X. Fan, H. Zhang, C. Leung, and C. Miao. 2016. Comparative study of machine learning algorithms for activity recognition with data sequence in home-like environment. In *2016 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*. 168–173.
[22] Hongqing Fang, Lei He, Hao Si, Peng Liu, and Xiaolei Xie. 2014. Human activity recognition based on feature selection in smart home using back-propagation algorithm. *ISA Transactions* 53, 5 (2014), 1629–1638.
[23] L. Ganong and M. Coleman. 2014. Qualitative research on family relationships. *Journal of Social and Personal Relationships* 31, 4 (2014), 451–459.
[24] Michelle L. Gattshall, Jo Ann Shoup, Julie A. Marshall, Lori A. Crane, and Paul A. Estabrooks. 2008. Validation of a survey instrument to assess home environments for physical activity and healthy eating in overweight children. *International Journal of Behavioral Nutrition and Physical Activity* 5, 1 (2008), 1–13.
[25] Jonathan Gershuny and Teresa Attracta Harms. 2016. Housework Now Takes Much Less Time: 85 Years of US Rural Women's Time Use. *Social Forces* 95, 2 (2016), 503–524.
[26] Zoubin Ghahramani. 2002. Hidden Markov Models. World Scientific Publishing Co., Inc., Chapter An Introduction to Hidden Markov Models and Bayesian Networks, 9–42.
[27] Kuo-Chung Hsu, Yi-Ting Chiang, Gu-Yang Lin, Ching-Hu Lu, Jane Yung-Jen Hsu, and Li-Chen Fu. 2010. *Strategies for Inference Mechanism of Conditional Random Fields for Multiple-Resident Activity Recognition in a Smart Home*. Springer Berlin Heidelberg, 417–426.
[28] Paul Kelly, Emma Thomas, Aiden Doherty, Teresa Harms, Órlaith Burke, Jonathan Gershuny, and Charlie Foster. 2015. Developing a Method to Test the Validity of 24 Hour Time Use Diaries Using Wearable Cameras: A Feasibility Pilot. *PLOS ONE* 10, 12 (12 2015), 1–15.
[29] Rebecca Killick and Idris A. Eckley. 2014. changepoint: An R package for changepoint analysis. *Journal of statistical software* 58, 3 (2014), 1 – 19.
[30] Rebecca Killick, P. Fearnhead, and Idris A. Eckley. 2012. Optimal Detection of Changepoints with a Linear Computational Cost. *J. Amer. Statist. Assoc.* 107, 500 (2012), 1590 – 1598.
[31] Ruppanner L. 2015. *Contemporary Family Issues*. Oxford University Press.
[32] O. D. Lara and M. A. Labrador. 2013. A Survey on Human Activity Recognition using Wearable Sensors. *IEEE Communications Surveys Tutorials* 15, 3 (2013), 1192–1209.
[33] V. I. Levenshtein. 1966. Binary codes capable of correcting deletions, insertions, and reversals. *Cybernetics and Control Theory* 10, 8 (1966), 707 – 710.
[34] Z. H. Lin and L. C. Fu. 2007. Multi-user Preference Model and Service Provision in a Smart Home Environment. In *2007 IEEE International Conference on Automation Science and Engineering*. 759–764.
[35] D. L'opez-G'omez and T. Sanchez-Criado. 2015. Analysing Hands-on-Tech Care Work in Telecare Installations. Frictional Encounters with Gerontechnological Designs. In *Aging and the Digital Life Course*, D. Prendergast and C. Garattini (Eds.). Berghahn, Chapter 9.
[36] M. Mort, C. Roberts, and B. Callen. 2013. Ageing with telecare: care or coercion in austerity? *Sociology of Health and Illness* 35, 6 (2013), 799–812.
[37] F. J. Ordó nez, G. Englebienne, P. de Toledo, T. van Kasteren, A. Sanchis, and B. Kröse. 2014. In-Home Activity Recognition: Bayesian Inference for Hidden Markov Models. *IEEE Pervasive Computing* 13, 3 (2014), 67–75.
[38] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, and others. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12, Oct (2011), 2825–2830.
[39] Dominique Picard. 1985. Testing and estimating change-points in time series. *Advances in applied probability* (1985), 841 – 867.
[40] James Pierce, Diane J. Schiano, and Eric Paulos. 2010. Home, Habits, and Energy: Examining Domestic Interactions and Energy Consumption. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. 1985–1994.
[41] Ronald Poppe. 2010. A survey on vision-based human action recognition. *Image and Vision Computing* 28, 6 (2010), 976 – 990.
[42] Markus Prossegger and Abdelhamid Bouchachia. 2014. *Multi-resident Activity Recognition Using Incremental Decision Trees*. Springer International Publishing, Cham, 182–191.
[43] P. Rashidi and D. J. Cook. 2009. Keeping the Resident in the Loop: Adapting the Smart Home to the User. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 39, 5 (2009), 949–959.
[44] Raúl Rojas. 2013. *Neural networks: a systematic introduction*. Springer Science & Business Media.
[45] P. Shannon-Baker. 2016. Making Paradigms Meaningful in Mixed Methods Research. *Journal of Mixed Methods Research* 10, 4 (2016), 319–334.
[46] Muhammad Shoaib, Stephan Bosch, Ozlem Durmaz Incel, Hans Scholten, and Paul J.M. Havinga. 2015. A Survey of Online Activity Recognition Using Mobile Phones. *Sensors* 15, 1 (2015), 2059–2085.
[47] David Silverman. 2006. *Interpreting Qualitative Data: Methods for Analyzing Talk, Text and Interaction (3rd edition)*. SAGE Publications.
[48] N. Sonck and H. Fernee. 2013. *Using smartphones in survey research: a multifunctional tool. Implementation of a time use app; a feasibility study.* The Netherlands Institute for Social Research.
[49] O. Sullivan and J. Gershuny. 2016. Change in Spousal Human Capital and Housework: A Longitudinal Analysis. *European Sociological Review* 32, 6 (2016), 864–880.
[50] Charles Sutton, Andrew McCallum, and others. 2012. An introduction to conditional random fields. *Foundations and Trends® in Machine Learning* 4, 4 (2012), 267–373.
[51] Tim van Kasteren, Athanasios Noulas, Gwenn Englebienne, and Ben Kröse. 2008. Accurate Activity Recognition in a Home Setting. In *Proceedings of the 10th International Conference on Ubiquitous Computing (UbiComp '08)*. 1–9.

[52] J. Wang, Z. Zhang, B. Li, S. Lee, and R. S. Sherratt. 2014. An enhanced fall detection system for elderly person monitoring using consumer home networks. *IEEE Transactions on Consumer Electronics* 60, 1 (2014), 23–29.

[53] Liang Wang, Tao Gu, Xianping Tao, Hanhua Chen, and Jian Lu. 2011. Recognizing multi-user activities using wearable sensors in a smart home. *Pervasive and Mobile Computing* 7, 3 (2011), 287 – 298.

[54] Simon J. Williams, Catherine Coveney, and Robert Meadows. 2015. 'M-apping' sleep? Trends and transformations in the digital age. *Sociology of Health & Illness* 37, 7 (2015), 1039–1054.

[55] Walter Zucchini and Iain L. MacDonald. 2009. *Hidden Markov Models for Time Series: An Introduction Using R.* CRC press.