Anomaly detection in elderly daily behavior in ambient sensing environments

Oya Aran[1], Dairazalia Sanchez-Cortes[1], Minh-Tri Do[1] [*], and Daniel Gatica-Perez[12]

[1] Idiap Research Institute, Martigny, Switzerland
[2] Ecole Polytechnique Federale de Lausanne (EPFL), Switzerland
`{oaran,dscortes,gatica}@idiap.ch`, `minhtrido@gmail.com`

Abstract. Current ubiquitous computing applications for smart homes aim to enhance people's daily living respecting age span. Among the target groups of people, elderly are a population eager for "choices for living arrangements", which would allow them to continue living in their homes but at the same time provide the health care they need. Given the growing elderly population, there is a need for statistical models able to capture the recurring patterns of daily activity life and reason based on this information. We present an analysis of real-life sensor data collected from 40 different households of elderly people, using motion, door and pressure sensors. Our objective is to automatically observe and model the daily behavior of the elderly and detect anomalies that could occur in the sensor data. For this purpose, we first introduce an abstraction layer to create a common ground for home sensor configurations. Next, we build a probabilistic spatio-temporal model to summarize daily behavior. Anomalies are then defined as significant changes from the learned behavioral model and detected using a cross-entropy measure. We have compared the detected anomalies with manually collected annotations and the results show that the presented approach is able to detect significant behavioral changes of the elderly.

Keywords: anomaly detection, healthcare, elderly care, sensor networks

1 Introduction

Pervasive and ubiquitous computing is essential to understand human behavior. Multimodal and more seamless embedded sensors support advances in human behavior understanding, covering a diversity of challenging topics including personality, emotions, human mobility and activities in everyday life [5,22,14,9,8]. The learned lessons make room to implement convenient applications aiming to enhance day to day living of people through their different age span.

The elderly constitute a group that deserves special attention to support their daily life [3]. Previous research has shown that seniors want to have choices

---

[*] Do is currently affiliated with Gameloft, Ho Chi Minh City, Vietnam.

for their living arrangements, i.e., preserving as long as possible their sense of personal space, independence and autonomy [6,30]. While nursing homes are a standing option, they have a great risk of overpopulation in the coming years [3] and in addition it may not be the ideal option for people who prefer to live independently. A good alternative is to consider converting traditional homes into low-cost smart homes. A comprehensive survey on such systems can be found in [4], which also discusses the design considerations like unobtrusiveness, scalability, energy efficiency, and security.

The smart home concept includes homes embedded with simple environmental sensors and more complex systems including audio, video and biometric systems. Systems that are easy to put in place, that do not demand maintenance from the user and are affordable in terms of cost, give elderly and their families the option to age at home while being monitored. There is evidence that opportunistic home surveillance prevents in some cases hospitalization [20]. While there are several significant advances on activity recognition in smart homes for elderly [23,12,10,31], relatively few studies tackle the challenging topic of detecting "unusual" behavior on elderly [28,25,21,13]. In most of these studies, experiments are performed monitoring either non-elderly subjects or very few elderly subjects, thus the proposed frameworks are not necessarily generalizable. Overall findings reveal that there is a compromise between high accuracy on anomalous behavior detection and subjects' privacy.

The raw information captured by the sensors can not be shared as such with the medical staff or used directly to detect changes in behavior automatically. Statistical models are needed to capture recurring patterns of daily life activities. This extracted knowledge about recurring patterns could be used to enrich the information displayed to the medical staff and improve the precision of early detections.

In this paper, we present a framework to analyze elderly daily behavior using only motion and state-change sensors. As an almost seamless and unobtrusive setting, it provides a promising approach for adoption among elderly and has shown to sufficiently capture day to day activities in real settings. The analysis used in this study is based on data gathered from 40 elderly homes on a four month period. In addition to the sensor data, there are annotations from questionnaires and daily activity journals. Our contributions in this paper can be summarized as follows.

1. We propose to use an abstraction layer, with the purpose of creating a common ground for different possible sensor configurations and sensor types. Given the high variability in the type, the number and the position of the sensors in different apartments, the abstraction layer enables an approach that could be generalized to different home environments and multiple datasets.

2. Based on the location and outing inferences on the abstraction layer, we propose a probabilistic behavior model to summarize daily activities of subjects from their sensor activation data. The probabilistic model takes into

account the location of the subject at each hour of the day and defines a likelihood of the subject's behavior based on her/his location and outings. This model, computed over a long period of sensor data, indicates where the subject spends her/his time as part of a daily routine.

3. We show that the presented behavior model can be used to detect anomalies by comparing the actual behavioral data of a subject with her/his usual behavior, as modeled by the behavior model. We follow an unsupervised approach and use a cross-entropy measure to indicate the predictability of the data, which is used as the anomaly indicator score. To detect whether there is an anomaly at a given time, the score is thresholded to obtain the anomaly detections.

In the next section, we present the related work in the literature. Section 3 presents our approach, the sensing environment, and the data used in this study. In Sections 4 - 6, we present our contributions on the sensor abstraction layer, the user behavioral model, and the anomaly detection. Section 7 provides discussion and conclusions.

2   Related work

The use of pervasive computing devices has recently supported advances in understanding personality, preferences, emotions, human mobility and daily activity routines from longitudinal data [5,22,14,9,8]. A significant population that would benefit from accurate inferences and clear understanding of daily activity routines is the elderly population. Systems able to provide opportunistic information to relatives or healthcare professionals in charge of elderly, would provide more confidence to elderly living independently.

In the context of smart home research, there have been centered efforts in inferring Activities of Daily Living (ADL) with state-of-the-art machine techniques using manually annotated data. Noury et al., [24] presented an attempt to analyse ADL using an elderly hospital environment. The scenario included a hospital suite embedded with infrared sensors used to record daily activity of an elderly woman for a two-month period. Their findings reported significant correlations between diurnal and nocturnal activities in elderly. Moreover, Hong et al., [15] reported 83.4% overall activity accuracy on seven daily activities. For the study, a three room apartment was equipped with 14 state-change sensors. The reported performance corresponds to a 28 days observational period from a 26 year old male.

More recently, [29] reported 91.3% accuracy inferring eight daily activities. A couple of subject houses were equipped with 10 wireless sensors (5 including ambient). The activity observations corresponded to 20 consecutive days of 10 healthy subjects (6 women, 4 men) between 28 and 79 years old. The activities

were annotated by the subjects with pen and paper, and also with a wireless device. Similarly, Pereira et al., reported up to 83% accuracy recognizing 7 typical movement activities. The framework combine sensing from wearable, portable and environmental sensors [27]. The movement activities were collected from 10 volunteers aged between 19 and 51, and later on manually annotated. Fleury et al.,[10] reported 13.7% global error rate on seven inferred activities. For the study, an in-lab smart home was equipped with six infrared sensors, four door contacts, one temperature sensor, eight microphones, 5 webcameras and a wearable kinematic sensor. Although results presented in [10] look promising, the design of the scenario including audio and visual sensors might not be in harmony with common elderly expectations regarding non-intrusive and privacy protection sensors at home.

It is worth to note that above cited works do not take under consideration the problem of anomaly detection. Anomaly detection "*refers to the problem of finding patterns in data that do not conform to expected behavior*" [7]. This topic is highly relevant in the context of ADL and sensor data in smart homes [18,17,21], and should be considered as critical for the deployment of elderly surveillance systems [16].

A Bayesian formulation is provided for anomaly detection in [26]. Behavioral patterns of the residents were extracted using Bayesian statistics, based on the raw measures of user activity, captured by several event-based sensors. The behavior was statistically estimated based on three probabilistic features: sensor activation likelihood, sensor sequence likelihood, and sensor event duration likelihood. The validation of the results has been performed on a data collected from three different home settings, with adult or elderly subjects living alone. The data is collected for 14 to 25 days.

Kim and Chung presented in [19] a framework that serves to monitor emergency situations of patients with chronic diseases. The framework proposes the use of wearable devices (including GPS), ambient and motion sensors as well as video cameras and speakers. The emergency module follows a rule-based semantic inference which outputs a list of actions to be taken. The module relies on the motion history image and continuous tracking.

The research in [28] proposes to compute a wellness index to capture abnormal behavior using several weeks of observations from elderly living alone. For the study, 6 wireless sensors (active/inactive) were installed in four elderly houses during several weeks. Nine ADL are inferred [28] from the embedded sensors and the wellness function aims to capture how "healthy" is the elderly person, i.e., being able to perform daily activities. The wellness function estimates maximum and minimum thresholds for the given activities computed after a trial run period of one week. After the trial period, warning messages are generated whenever wellness values passed the thresholds. More recently, [11] approached the inference of abnormalities and detection of changes in routine behavior. For the study, three months data of 10 users was synthetically generated, the data

consisting of vital signs (heart rate, blood pressure, blood sugar, respiration and temperature), location, activity and lifestyle. With 8 activities, the accuracy of activity sequence is estimated above 87% with normal observations and above 90% with abnormal observations. After the abnormalities are detected, they propose a method that uses fuzzy rules to describe actions to be taken.

In the above studies, the following observations can be made on the data used for the analysis.

- The data comes from few households: It is challenging to install sensors on different households. However, it is also challenging to analyze data coming from multiple households, in which there are different sensor configurations due to different floor plans, number of rooms, etc. In this paper, we propose to use an abstraction layer, which allows us to analyze the data coming from different sensor configurations. The data that we analyze come from 40 different households.

- The data is collected from non-elderly people: It is a challenging task to collect data from elderly people due to privacy and technology adoption reasons. Due to this, most studies use data coming from adults and young adults. In our study, we use data coming from subjects aged 62 to 96, with an average age of 84, spanning a truly elderly population.

- The data hardly contains ground-truth labels: Annotating daily routine behavior is cumbersome. In particular for elderly people, they have troubles remembering to annotate during their daily routines and have to use the classical pen and paper interface for the annotations as they do not always cope well with technological devices such as smart phones or tablets. These factors limit the amount of ground truth labels that can be obtained. In our study, we are also limited by the amount of available annotations. We use manual annotations of ADL coming from one of the subjects, spanning 8 days.

3   Our Approach

The overall objective of this study is to analyze daily routine behavior of elderly people in their apartment through ambient sensors. In particular, we aim to detect anomalies and significant changes in the behavior, with the assumption that these anomalies may signal health related problems. In the next sections, we present the sensing environment and the details of the dataset that has been used in the study.

3.1   Sensing Environment

We have used data coming from a commercial product, which includes ambient sensors installed in apartments where elderly live [1]. The product enables the

caregivers to monitor the patients' activity and daily behavior based on the sensor data. The sensor system includes a base unit and wireless, battery-powered sensors which are installed in strategic locations around the apartment. The sensors include open/close door (for main entrance and fridge), force sensors (for bed and chair), motion/activity sensors in the living room, bedroom, etc., and smoke sensors.

3.2   Dataset

The data used for experiments was collected in the context of a Swiss project from mid-December 2013 to early April 2014. The data collection was led by DomoSafety [1] and La Source, School of Nursing Sciences, University of Applied Sciences of Western Switzerland [2]. The collected data includes 45 unique subjects, covering a total of 104 days. During this period, some of the subjects dropped out from the data collection due to various reasons. For further analysis and to maintain consistency among the available recorded dates, five subjects were discarded from the study. All data used in the paper was handled in an anonymized way.

The recruitment of the subjects lasted a couple of months before the start of the data collection. For each recruited subject, an initial questionnaire was applied at the time of the recruitment, prior to the data collection. The average age of the subjects is 84.3 years, with a minimum age of 62 years and a maximum of 96. The 40 subjects in the study comprise 27 females and 13 males.

Regarding external help for daily living, 31 subjects declared receiving professional cleaning services, 25 using food delivery services, and 28 subjects declared receiving help from family. 25 subjects reported climbing stairs without help. 35 subjects declared having regular outings. 29 subjects declared to be widow/er, 3 married, 4 single and 2 divorced. Other questions also captured general health problems, such as medicines taken, nutritional status, overall physical and functional state.

Sensor data   The original data comprises several measurements at different frame rates from various types of sensors. The number, types, and location of sensors vary for each installation depending on the configuration of the corresponding apartment. Figure 1 shows an example configuration of the analyzed static sensors from one subject's apartment.

Journal annotations   With the aim of having a source of ground truth, the collection includes daily activity journal (DAJ) annotations that has been performed twice during the study. A first set of annotations was collected at the start of the study from three subjects, with the purpose of evaluating the newly installed sensors. A second and a more reliable set of annotations was collected
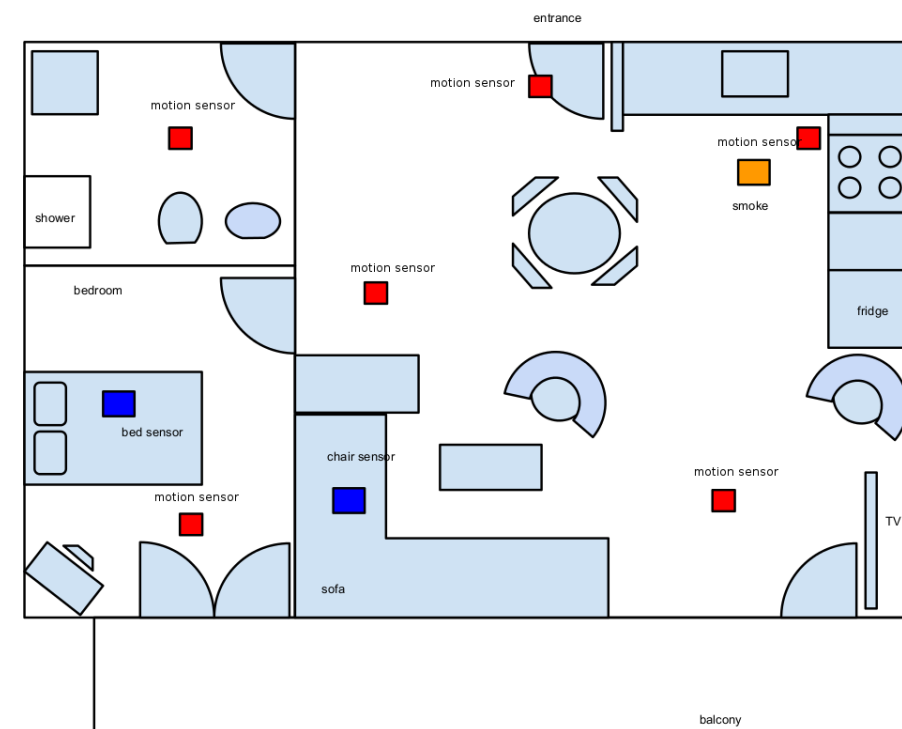
Fig. 1. Floor plan with static sensors from one subject's apartment.

from a single subject towards the end of the data collection. The resident was asked to report the start and end time of the following activities: *bathroom visits*, *meals*, *visits*, and *outings*. To make the annotation task easier for the subjects, the diaries were split based on activities performed in each location, printed in paper, and placed accordingly in the apartment.

4 Sensor Data Abstraction layer

The sensor network used in our study continuously captures a set of measurements at different frame rates from various types of sensors. The number, types, and location of these sensors vary in different installations depending on the configuration of the corresponding apartment. For instance, a two-room apartment will be equipped with less sensors than a five-room one. A large living room requires more than one activity sensor to cover its volume.

We built an abstraction layer to overcome the variability of the sensor configuration. This abstraction layer takes all the sensor values as input and generates

---

Algorithm 1 Location Inference

1: Input: activity signals $\mathbf{s} = (s_i.time, s_i.location)_{i=1..n}$, location set $\mathcal{L}$
2: Output: sequence of location events $\mathbf{l} = (l_i.start, l_i.end, l_i.location)_{i=1..m}$
3: Method: update location $\mathbf{l}$ based on the duration of the activity signals, $\mathbf{s}$

---

sequences of *events* as output. An *event* is simply defined by its respective start and end timestamps and a label. We consider two event types in this study: **Locations** and **Outings**.

This abstraction layer produces a new representation of the data that simplifies the raw sensor measurements and allows the system to be independent of the number and types of sensors found in an apartment. In the next sections, we describe the modules which extract abstract events from the raw sensor data. We then evaluate its performance by comparing the sequences of events estimated by the abstraction layer to a written diary that serves as ground truth (see Section 4.3).

4.1   Location inference

In the study, the sensors are installed in different positions of the apartment. When a sensor fires, the location of the resident is registered. We developed an algorithm that makes use of the localized activity signals to infer the actual location of the resident. Given the activity signals, the algorithm outputs a sequence of location events, consisting of the start time, end time, and the location label.

The algorithm keeps track of the location by examining the set of incoming signals and updates the location state accordingly. We assume that the resident does not change the location when no activity has been detected. In other words, the location is updated only if one or several activity signals occur. In practice, we can guarantee this assumption by installing the sensors at key locations, to make sure that the system can detect activity whenever the resident enters a new location.

At a given time, if there is exactly one activity signal, the location update is straightforward. However, it might happen that multiple signals arrive at the same time, and that these signals do not come from the same location. This issue occurs when the coverage area of different sensors are overlapping (e.g., a sensor inside the bathroom and a sensor in front of the bathroom door can both detect activity when the resident enters the bathroom) or when the sensors are not perfectly synchronized. In this situation, the algorithm needs to choose one location among several candidates. Algorithm 1 shows the summary of the location inference algorithm.

Algorithm 2 Outing classification

1: Input: raw signals **s**, indices of two consecutive entrance door signals k, h
2: Output: Indicate if the period $(s_k.time, s_h.time)$ corresponds to an outing
3: Method: Classify as outing if there are no activities in between two consecutive entrance door signals k, h

---

4.2 Outing detection

To analyze the subject's behavior and make sense of sensor data, it is important to know whether the subject is inside the apartment. In this section, we present an algorithm to infer outing events by using information from the entrance door and the activity in the apartment. The door openings provide potential outing event candidates, which are then verified by checking if there is any activity inside the apartment during the period of interest.

We assume that the resident always closes the door after leaving the apartment and opens the entrance door to enter the apartment. This assumption simplifies the problem as we only need to check if a period between two consecutive door signals is actually an outing. The outing classification decision is made after verifying that there are no activities inside the apartment between two consecutive door signals. Algorithm 2 shows the summary of the outing classification algorithm.

4.3 Evaluation

We evaluate the quality of the outputs of the abstraction layer with a daily activity journal (DAJ) filled by one subject for 8 consecutive days. We use the bathroom and outing annotations from the DAJ as ground truth data to evaluate the performance of the abstraction layer on predicting the bathroom visits and outings events.

To analyze the sensor data, we split the time scale in five-minute segments and consider each segment as the basic unit in our evaluation. For each of the events we consider (outings and bathroom usage), both for the journal annotations and sensor data, if there is any reported event within a given segment, that segment is labeled as true for the corresponding event. For the evaluation, from these time-aligned binary values, we compute the confusion matrix between two systems: the ground truth values coming from the journal annotations and the sensor values. We consider the journal annotations as the reference.

Location: To evaluate location inference, we focus on the bathroom events, as we have ground truth annotations for the bathroom usage. The abstraction layer provides a sequence of bathroom events that we compare to the DAJ

ground truth data. A total of 52 bathroom visits have been reported by the resident, which corresponds to 94 segments of 5 minutes. 13 segments have been wrongly labeled as non-bathroom events by the system, and 46 segments have been wrongly labeled as bathroom visits. This corresponds to a precision of 64% and a recall of 86% for the bathroom usage. Table 1(a) shows the confusion matrix for the bathroom events.

Outings: We correctly predict 245 segments (20.4 hours) out of 21.8 hours of outings reported in the diary. A total of 39 segments (3.25 hours) are wrongly assigned to outings. Table 1(b) shows the confusion matrix for the outing events. The precision is 86% and the recall is 94%.

Table 1. Abstraction layer evaluation based on the comparison of the diary ground truth data and the sensor data.. (a) Confusion matrix of the bathroom usage events, and (b) Confusion matrix of the outing events.

(a)

	Sensor-False	Sensor-True
Journal-False	2164 (TN)	46 (FP)
Journal-True	13 (FN)	81 (TP)

(b)

|	Sensor-False	Sensor-True
Journal-False	2003 (TN)	39 (FP)
Journal-True	17 (FN)	245 (TP)

5   User behavioral model

5.1   Location based probabilistic model

We introduce a statistical model for summarizing behavioral data of a subject. For each subject, we assume that the location at a given timestamp depends only on the hour of the day. Let $\mathbf{l} = \{l_t\}$ be the location sequence and $\mathbf{h} = \{h_t\}$ denotes the sequence of hour-of-the-day where $h_t \in \{1..24\}$. The likelihood of the data can be defined as follows:

$$p(\mathbf{l}; \mathbf{h}, \theta) = \prod_h \prod_l \theta_{h,l}^{n(l,h,\mathbf{l},\mathbf{h})}$$

where $\theta_{h,l} = P(l|h)$ denote the probability of being at location l at *hour* h, $n(l, h, \mathbf{l}, \mathbf{h}) = \sum_i \mathbb{1}(l_i = l \wedge h_i = h)$ is the count of location l in time slot h.
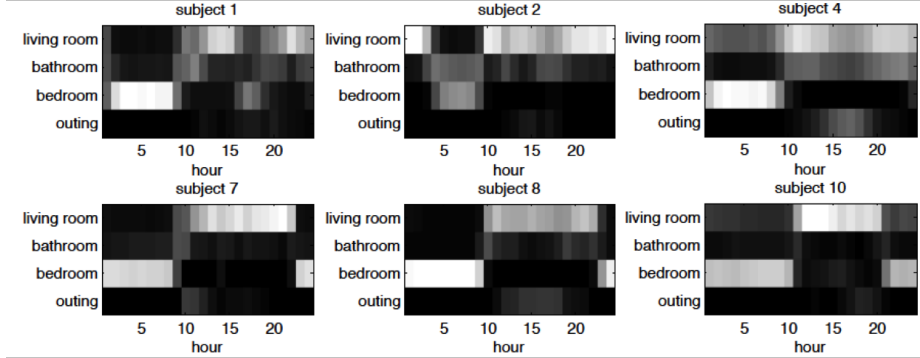
Fig. 2. Samples of user behavioral model. Brightness of each cell represents the conditional probability of being at a location at a given hour.

Note that we treat **h** as external information instead of the observation in the probabilistic model.

Let $(\mathbf{l}_s, \mathbf{h}_s)$ be the data for subject s, the parameter of the categorical distribution is computed as follows:

$$\theta_{h,l}^{(s)} = \frac{n(l, h, \mathbf{l}_s, \mathbf{h}_s) + \alpha}{\sum_l n(l, h, \mathbf{l}_s, \mathbf{h}_s) + \alpha}$$

where the scalar $\alpha > 0$ is introduced for the regularization purposes. In our experiment, we see that the results do not change significantly with respect to α, thus we set $\alpha = 1$.

Figure 2 illustrates different samples of behavioral model, which summarize daily activities of different subjects. For each hour of the day, we use a discrete distribution over the four location categories to characterize the mobility pattern of the subject.

5.2 Discovering common behavioral patterns

The model reveals several behavioral patterns such as the going to bed time, waking up time, or sleep interruptions during the night. As an analysis of the common behavioral patterns shared by different subjects, we applied clustering on the behavioral profiles of different users. We used k-means clustering with two clusters. For each subject, we concatenated values of the behavioral patterns of that subject into a vector, forming a feature vector of size 96 (24 hours x 4 locations). Figure 3 shows the mean profiles in each cluster. We see that there are two typical behaviors among the participants of the study. A majority of the
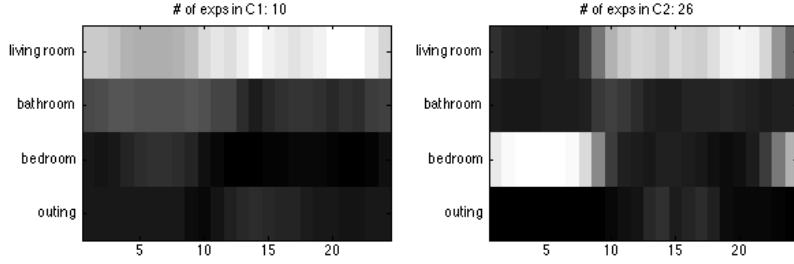
Fig. 3. Mean behavior profiles for two clusters.

subjects spend their time in the living room and sleep in their bedroom at night, with occasional outings and several bathroom visits during the day (Figure 3, right). However, the other group of subjects does not seem to use their bedroom for their night sleep (Figure 3, left). This is an important observation, as it has implications particularly on the placement of the sensors, i.e. the bed sensor.

6   Anomaly Indicator

We define anomalies as events, which are different than the subjects' past behavior, as modeled by the behavior model defined in Section 5. There can be several reasons for anomalies, ranging from sensor failures to particular health problems that the patient might have. Accurate and timely detection of anomalies are highly informative for the caregivers and can be life saving in particular situations.

6.1   Measuring the data predictability

We assume that any anomaly can be measured as deviations from the routine behavior and define a cross-entropy measure as an indicator for the predictability of the data. Given the data $\mathbf{l}, \mathbf{h}$ and the learned behavioral model θ, the cross-entropy is computed as follows:

$$H(\mathbf{l}, \theta; \mathbf{h}) = - \sum_{t=1..T} \frac{1}{T} log_2 p(l_t|\theta, h_t) = - \sum_{t=1..T} \frac{1}{T} log_2 \theta_{h_t, l_t} \tag{1}$$

The cross-entropy measures the average number of bits to encode the data given the learned behavioral model. A low entropy indicates that the empirical distribution is well predicted by the learned model, while a high entropy indicates that the model does not accurately predict the data.

To simulate an online setting, we divide the data of each subject into one-week periods. At the ith week, we learn the model from the behavioral data of the $(i-1)$ previous weeks, and estimate the cross-entropy for each hour and each day of the ith week. Figure 4 illustrates a few samples of our predictability indicator based on cross-entropy. The first line shows the anomaly scores estimated per hour of the day. The second line shows the scores estimated per day. The third line shows the mobility pattern of the subjects including the locations inside the apartment and the outings. Note that for the first week, the training data is empty and uniform distribution is used by default. In many cases, the predictability increases quickly after a few weeks if the daily behavioral data is repetitive, until there are some anomaly in mobility pattern. For example, the plot of subject 8 shows that the first month of data is highly predictable. Then, the fact that the mobility pattern has changed as less repetitive (outing several hours depending on the day) is reflected by the increase in the number of bits needed to encode the new data. Using a shorter temporal resolution, as for the per hour of the day estimation, allows for more rapid estimations. It can be seen that the changes in the behavioral patterns are captured as quickly as possible, immediately following the pattern change, allowing for a shorter response time.

Overall, we can visually find a correlation between the anomaly indicator and the change in mobility behavioral data. This means that the proposed indicator can be potentially used for anomaly detection, with an appropriate threshold value. In the next section, we investigate and compare the accuracy of anomaly detection per day and per hour of the day, when used with different thresholds.

6.2   Anomaly Detection Evaluation

For the validation of the anomaly detector, we need ground truth values that indicate various types of anomalies that could occur. One type of anomaly could be due to sensor malfunction. Either a sensor stops responding or starts providing faulty readings. For example, in Figure 4, for each subject, there are one or more days, where the subject's location is always recorded as bedroom (green lines), which most probably reflects sensor failures. However, the sensor failures in this dataset have not been logged, thus we do not have access to that information. Another type of anomaly could be due to a significant change in the behavior of the subject, which could signal a health related condition. For example, the subject starts spending more and more time in bed, showing a decrease in the number of outings or the use of the kitchen, etc. If available, subject's health information could also be used as ground truth, in case the person had any important health conditions. Unfortunately, in the dataset used for this study, such information does not exist. To compensate for the lack of ground truth, we have collected manual annotations of anomalies. The location mappings of the subjects is shown to the annotators (i.e. such as the ones in the third row of Figure4) and based on this information the annotators indicated the days that they think could be an anomaly. The manual annotations indicate only the
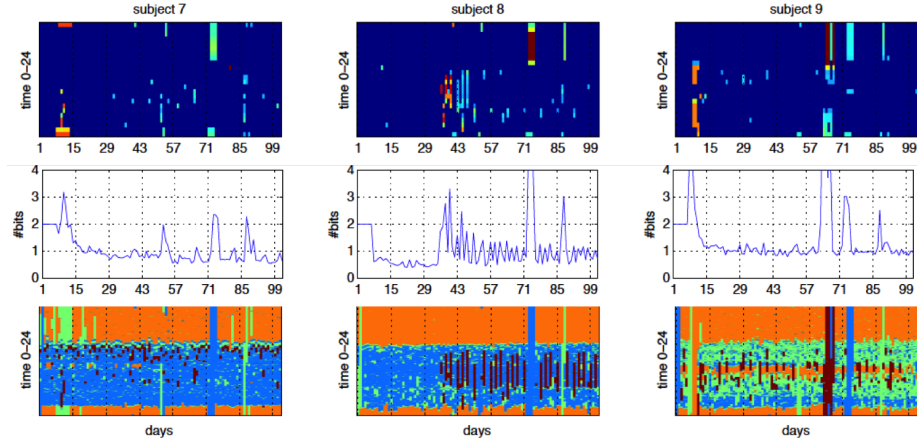
Fig. 4. Anomaly scores on three sample subjects. The figures in the first, second and third rows correspond to scores per hour of the day, scores per day, and the mobility pattern, respectively. For the per hour of the day results in the first row, the anomaly score increases from dark blue to dark red. For the mobility pattern, blue corresponds to living room, green corresponds to bathroom, orange corresponds to bedroom, and red corresponds to outings.

day of the anomaly as we decided that any shorter duration could have been harder and unreliable to manually annotate with the information shown during annotation.

The comparison of the anomaly detections with the manual annotations is done as follows:

- The detected anomaly scores are thresholded with a set threshold.

- The hourly detections has been mapped such that if there is an anomaly score higher than the threshold in any hour of the day, that day is considered as an anomaly.

- If a detected anomaly is on the same day as one of the manually annotated anomalies, it is considered as a correct detection, otherwise it is considered as a false detection.

The evaluation is done on a sub-dataset to filter out missing data and considering only subjects who live alone. The final dataset used in the evaluation corresponds to 104 days from 36 subjects. Figure 5 shows the Receiver Operating Characteristic (ROC) curves for anomaly detection. We have used the quantile values to set a threshold and the figure shows the ROC curves drawn with thresholds correspond to the quantile of the data in the range of $[0.10 - 0.995]$. We observe that the two approaches give similar results, with the hourly detection giving slightly better results at the threshold levels determined by 0.975 and 0.95 quan-
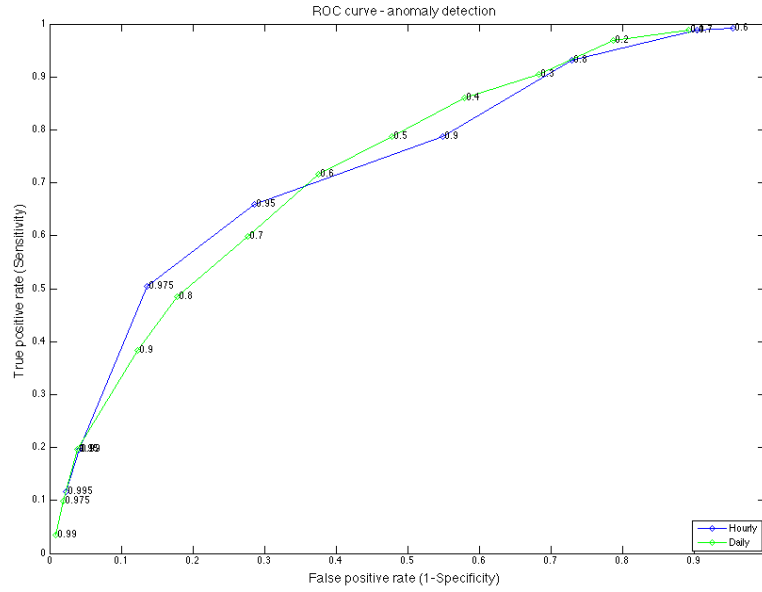
Fig. 5. ROC curves for anomaly detection for two different approaches, per day and per hour of the day. The quantile values used to determine the thresholds are shown next to each data point.

tile. For a threshold on the 0.95 quantile of anomaly scores detected every hour of the day, the true positive rate is 0.66 and the false positive rate is 0.29. Using per day detection, 0.72 true positive rate can be obtained with a higher false positive rate, 0.38, based on a threshold determined by the 0.60 quantile of the data.

In addition to the performance measures, the hourly detection has the advantage of reporting the anomaly earlier, at the latest one hour after the anomaly whereas the daily detection needs to wait until the end of the day. From this perspective, shorter temporal resolutions are preferable over the longer ones given similar performance.

7 Discussion and Conclusion

We presented an analysis of real-life sensor data collected from a large number of households with elderly subjects. Our framework first defines an abstract layer to create a common ground for different sensor configurations. We aim to build a behavior model, which represents the subjects' daily behavior based on the

events extracted from the abstraction layer. Based on the location and outings events in the abstract layer, our probabilistic behavior model captures the repetitive daily routines of the subjects. The use of an abstraction layer facilitates to generalize the sensor information in terms of events. Thus, the created events are independent of the sensor configuration and type, and summarize the activity of the subjects. For example, in order to detect the living room location in a big apartment, it may be necessary to install two sensors in the living room in order to cover the whole space, whereas in a small apartment, one sensor would be enough. When converted to events, regardless of the number of sensors in the living room, there will be one single event of being in the living room. This generalization allows us to use the same approach in different households with different configurations. While our framework is general, we focus on building a model of location data, which are considered as key in many ubiquitous applications. However, the abstraction approach is not limited to locations and outings but can also be formulated for activity level, event sequence, etc., which would provide additional information to model the subject's behavior. The integration of other data types can be theoretically done by adding more variables to the system.

Once the behavior model is created, we use this model to detect anomalies. We propose an indicator of predictability based on the cross-entropy measure for the detection of anomalies. Our method compares the new observed behavioral data with the learned distribution to measure how well the learned behavioral model predicts the new data. By using cross-entropy, we define the anomaly indicator as the average number of bits to encode the new data. This normalized measurement facilitates the comparison between data periods, subjects, and also future behavior models with additional data types.

One of the assumptions that we used is that there is only one subject in the apartment. Having this assumption, any visitors in the apartment could be detected as an anomaly, unless they are regular visitors. The regular visitors, such as the cleaning person, who come at fixed times in a week and following a similar routine in the apartment, could be captured by the behavior model. However, this has not been confirmed in the current study as we did not have any accurate information on the visitors of the subjects.

The overall aim in anomaly detection is to detect any type of anomaly and inform the caregivers about the anomaly as soon as it is detected. If it is a sensor anomaly, the sensor needs to be changed as soon as possible. If it is a change in the daily routine of the subject, a visit by the caregiver would be necessary. If it is a sudden and persistent change, it may be an emergency situation. In this study, we only looked at the detection of anomalies without investigating the type of anomaly. To further validate the performance of the anomaly detection, there is a need to gather real-life ground truth labels that indicate different types of anomalies, in particular, the health related ones. The availability of ground truth information for different type of anomalies would also enable us to develop methods that could identify the type of anomaly at the time of detection, which

we leave as part of future work. It is also important to note that the detection should be as accurate as possible so that the caregivers are provided with precise information. The accuracy of anomaly detection is dependent on the detection approach as well as the behavior model, which are both based on the sensor data and the abstraction layer. We observe that most of the errors stem from faulty sensor data or wrong sensor placement. Improving sensor quality and a more careful placement of sensors will automatically lead to higher accuracy.

References

1. DomoSafety. `http://www.domo-safety.com/`
2. Institut et Haute Ecole de la Sante La Source. `http://www.ecolelasource.ch`
3. United Nations, population division. `http://www.un.org/esa/population/publications/worldageing19502050/pdf/80chapterii.pdf`
4. Alemdar, H., Ersoy, C.: Wireless sensor networks for healthcare: A survey. Computer Networks 54(15), 2688–2710 (2010)
5. Alshamsi, A., Pianesi, F., Lepri, B., Pentland, A., Rahwan, I.: Network diversity and affect dynamics: The role of personality traits. PloS one 11(4), e0152358 (2016)
6. Borsch-Supan, A., Hajivassiliou, V., Kotlikoff, L.J.: Health, children, and elderly living arrangements: A multiperiod-multinomial probit model with unobserved heterogeneity and autocorrelated errors. In: Topics in the Economics of Aging, pp. 79–108. University of Chicago Press (1992)
7. Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: A survey. ACM Comput. Surv. 41(3), 15:1–15:58 (Jul 2009), `http://doi.acm.org/10.1145/1541880.1541882`
8. Do, T.M.T., Dousse, O., Miettinen, M., Gatica-Perez, D.: A probabilistic kernel method for human mobility prediction with smartphones. Pervasive and Mobile Computing 20, 13–28 (2015)
9. Farrahi, K., Gatica-Perez, D.: What did you do today?: discovering daily routines from large-scale mobile data. In: Proceedings of the 16th ACM international conference on Multimedia. pp. 849–852. ACM (2008)
10. Fleury, A., Noury, N., Vacher, M.: Improving supervised classification of activities of daily living using prior knowledge. Digital Advances in Medicine, E-Health, and Communication Technologies p. 131 (2013)
11. Forkan, A.R.M., Khalil, I., Tari, Z., Foufou, S., Bouras, A.: A context-aware approach for long-term behavioural change detection and abnormality prediction in ambient assisted living. Pattern Recognition 48(3), 628 – 641 (2015), `http://www.sciencedirect.com/science/article/pii/S0031320314002660`

12. Foroughi, H., Aski, B.S., Pourreza, H.: Intelligent video surveillance for monitoring fall detection of elderly in home environments. In: Computer and Information Technology, 2008. ICCIT 2008. 11th International Conference on. pp. 219–224. IEEE (2008)
13. Franco, C., Demongeot, J., Villemazet, C., Vuillerme, N.: Behavioral telemonitoring of the elderly at home: Detection of nycthemeral rhythms drifts from location data. In: Advanced Information Networking and Applications Workshops (WAINA), 2010 IEEE 24th International Conference on. pp. 759–766. IEEE (2010)
14. Golder, S.A., Macy, M.W.: Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. Science 333(6051), 1878–1881 (2011)
15. Hong, X., Nugent, C.D.: Segmenting sensor data for activity monitoring in smart environments. Personal and ubiquitous computing 17(3), 545–559 (2013)
16. Hsu, H.H., Chen, C.C.: Rfid-based human behavior modeling and anomaly detection for elderly care. Mobile Information Systems 6(4), 341–354 (2010)
17. Jakkula, V.R., Cook, D.J.: Detecting anomalous sensor events in smart home data for enhancing the living experience. Artificial intelligence and smarter living 11(201), 1 (2011)
18. Jakkula, V.R., Crandall, A.S., Cook, D.J.: Enhancing anomaly detection using temporal pattern discovery. In: Advanced intelligent environments, pp. 175–194. Springer (2009)
19. Kim, S.H., Chung, K.: Emergency situation monitoring service using context motion tracking of chronic disease patients. Cluster Computing 18(2), 747–759 (2015), http://dx.doi.org/10.1007/s10586-015-0440-1
20. Kornowski, R., Zeeli, D., Averbuch, M., Finkelstein, A., Schwartz, D., Moshkovitz, M., Weinreb, B., Hershkovitz, R., Eyal, D., Miller, M., et al.: Intensive home-care surveillance prevents hospitalization and improves morbidity rates among elderly patients with severe congestive heart failure. American heart journal 129(4), 762–766 (1995)
21. Lotfi, A., Langensiepen, C., Mahmoud, S.M., Akhlaghinia, M.J.: Smart homes for the elderly dementia sufferers: identification and prediction of abnormal behaviour. Journal of ambient intelligence and humanized computing 3(3), 205–218 (2012)
22. Madan, A., Cebrian, M., Lazer, D., Pentland, A.: Social sensing for epidemiological behavior change. In: Proceedings of the 12th ACM international conference on Ubiquitous computing. pp. 291–300. ACM (2010)
23. Nasution, A.H., Emmanuel, S.: Intelligent video surveillance for monitoring elderly in home environments. In: Multimedia Signal Processing, 2007. MMSP 2007. IEEE 9th Workshop on. pp. 203–206. IEEE (2007)
24. Noury, N., Hadidi, T., Laila, M., Fleury, A., Villemazet, C., Rialle, V., Franco, A.: Level of activity, night and day alternation, and well being measured in a smart hospital suite. In: Proc. IEEE–EMBC. vol. 8, pp. 20–24 (2008)
25. Novák, M., Biňas, M., Jakab, F.: Unobtrusive anomaly detection in presence of elderly in a smart-home environment. In: ELEKTRO, 2012. pp. 341–344. IEEE (2012)
26. Ordóñez, F.J., Toledo, P., Sanchis, A.: Sensor-based bayesian detection of anomalous living patterns in a home setting. Personal Ubiquitous Comput. 19(2), 259–270 (Feb 2015), http://dx.doi.org/10.1007/s00779-014-0820-1
27. Pereira, J.D., da Silva e Silva, F.J., Coutinho, L.R., de Tácio Pereira Gomes, B., Endler, M.: A movement activity recognition pervasive system for patient monitoring in ambient assisted living. In: Proceedings of the 31st Annual ACM Symposium on Applied Computing. pp. 155–161. ACM (2016)

28. Suryadevara, N.K., Mukhopadhyay, S.C.: Wireless sensor network based home monitoring system for wellness determination of elderly. IEEE Sensors Journal 12(6), 1965–1972 (2012)
29. Urwyler, P., Rampa, L., Stucki, R., Büchler, M., Müri, R., Mosimann, U.P., Nef, T.: Recognition of activities of daily living in healthy subjects using two ad-hoc classifiers. Biomedical engineering online 14(1), 1 (2015)
30. Wiles, J.L., Leibing, A., Guberman, N., Reeve, J., Allen, R.E.S.: The meaning of "ageing in place" to older people. The Gerontologist 52(3), 357–366 (2012)
31. Yoo, J.H., Ko, J.G., Chung, Y.S., Jung, S.U., Kim, K.H., Moon, K.Y., Chung, K.: Design of embedded multimodal biometric systems. In: Signal-Image Technologies and Internet-Based System, 2007. SITIS'07. Third International IEEE Conference on. pp. 1058–1062. IEEE (2007)