ORIGINAL ARTICLE

Sensor-based Bayesian detection of anomalous living patterns in a home setting

Fco Javier Ordóñez · Paula de Toledo · Araceli Sanchis

Abstract In this paper, we present an automated behavior analysis system developed to assist the elderly and individuals with disabilities who live alone, by learning and predicting standard behaviors to improve the efficiency of their healthcare. Established behavioral patterns have been recorded using wireless sensor networks composed by several event-based sensors that captured raw measures of the actions of each user. Using these data, behavioral patterns of the residents were extracted using Bayesian statistics. The behavior was statistically estimated based on three probabilistic features we introduce, namely sensor activation likelihood, sensor sequence likelihood, and sensor event duration likelihood. Real data obtained from different home environments were used to verify the proposed method in the individual analysis. The results suggest that the monitoring system can be used to detect anomalous behavior signs which could reflect changes in health status of the user, thus offering an opportunity to intervene if required.

Keywords Behavior patterns · Bayesian statistics · Anomaly detection · Wireless sensor networks

F. J. Ordóñez (✉) · P. de Toledo · A. Sanchis
Computer Science Department, University Carlos III of Madrid,
Leganés, Madrid, Spain
e-mail: fordonez@inf.uc3m.es

P. de Toledo
e-mail: mtoledo@inf.uc3m.es

A. Sanchis
e-mail: masm@inf.uc3m.es

1 Introduction

Population aging is currently having a significant impact on health care systems [25]. Improvements in medical care are contributing to increase the survival rate among the elderly; thus, cognitive impairments associated with aging will increase [14]. It has been estimated that one billion people will be over the age of 60 by the year 2025 [22]. As demographics age and the burden of healthcare on society increase, the need for finding more effective ways of providing care and support to the disabled and elderly at home becomes more predominant. Automated and ambient systems for anomalous behavior detection serve a dual role [11]: (1) to increase the safety and the sense of security of people living on their own, and (2) to allow elderly patients to be self-reliant longer, fostering their autonomy.

Among the available technologies for pervasive activity monitoring, wireless sensor networks (WSNs) are considered one of the most promising approaches, due to their suitability to supply constant supervision, flexibility, low cost, and rapid deployment [3]. Besides, the inherent non-intrusive characteristics of these networks have been proved to suit perfectly with environments where privacy and user acceptance are required [7]. If a smart house can be instrumented with one of these networks, the occupants would have a better chance to live safely and independently [13], especially when they suffer from chronic and disabling conditions, such as Parkinson's or Alzheimer's disease. Several studies have employed these technologies to monitor the activities of the individuals at home, in order to predict daily behavior [6, 16] and eventually launch alarms to alert relatives, caregivers, or healthcare personnel when inadequate behaviors are detected [21].

The algorithm that we propose in this paper is aimed to statistically identify anomalous human behavioral patterns,

a problem already present in the literature which has been addressed by several authors. Recently, Aztiria et al. [1] developed an algorithm that compares the behavior of a user with a set of previously discovered frequent behaviors to identify possible shifts. Employing a set of atomic actions, the authors defined a likelihood value for the current behavior of the user. The number of modifications required to turn the current behavior into a frequent behavior is used as a metric to classify a sequence of actions as an anomaly.

Shin et al. [15] developed an abnormality detection system that employed several features about the activity and mobility of the users for identifying anomalous behavioral patterns. They proposed a system designed to monitor the daily life of the elderly in their home using several IR motion sensors. In order to represent the behavior of the users, they employed three different features, namely activity level, mobility level, and non-response interval. A one-class classification method based on a support vector data description (SVDD) algorithm was employed to combine the information provided by these features and to classify normal behavior patterns. They used both real and simulation data to validate the proposed abnormality detection algorithm in real and possible anomalous situations. This approach yielded 74.2 % sensitivity and 85.8 % specificity for real data and a higher performance in the simulation data.

Yin et al. [24] also used a one-class support vector machine (SVM) approach to detect anomalous activities of a user from body-worn sensors. Their approach follows a two-phase pipelined process, which combines a SVM for filtering out the normal activities with a collection of secondary classifiers for anomalous activity detection, aimed at reducing false-positive rate.

Another approach based on a SVM was proposed by Palaniappan et al. [12]. In this case, the authors employed a multiclass SVM for recognizing the normal activities, and the anomalous activities were detected by ruling out all possible activities that could be performed from the current activity. The whole system was focused on identifying anomalous activities with less computational time, in order to work efficiently in real time. Indeed, SVMs have been used for anomalous behavior detection not only with sensor data, but also with video devices [23]. Wu et al. defined a method where a discrete Fourier transformation was used to obtain different features from video data, and a SVM was then employed to classify the instances into normal or anomalous behaviors.

Other approaches to model behavioral patterns are based on circadian activity rhythms (CARs). Virone et al. [19] developed a monitoring system focused on elderly people living alone, which established behavioral patterns using a statistical predictive algorithm that modeled CARs and their deviations. They employed a system composed of different types of wireless sensors to monitor the health status and the behavior of an individual within his living environment [18]. The system analyzed different statistical metrics, employing the activity and presence level of the user to detect deviations in the behavioral patterns. They also constructed a credible interval to identify behaviors considered as non-habitual.

However, modeling human behavior is still a challenging problem because there are changes in how each person performs a specific activity and the human actions are usually executed in a non-deterministic way. In order to handle such issues, Bayesian methods provide a good framework to develop modeling techniques that specifically take uncertainty into account. Indeed, Bayesian statistics has been proved to be particularly well suited for automatic analysis of human behavior [4] and anomaly detection [5]. In this paper, we postulate that we can provide an improved estimation of the behavior of a person statistically through Bayesian probabilistic modeling. In order to accomplish that, we propose three probabilistic features related to the monitoring system, that are not present in the literature, to the best of our knowledge.

2 Sensing system and data collection

In this paper, we have employed datasets generated by a set of simple state-change sensors installed in three different environments. The sensing system was chosen according to two main criteria: ease of installation and minimal intrusion. Sensors that need to be worn on the body may be considered intrusive by the user, and sensors that are easy to install can increase the overall acceptance of the system.

The system consists of a set of wireless network nodes to which simple binary sensors are attached. Such wireless network was manufactured by RFM and included an energy efficient proprietary firmware. Each wireless node can be used with a large variety of sensors. We used this technology because it offers a simple and cheap way to deploy a highly scalable monitoring system. In our case, a collection of various sensors to measure different situations were employed: pressure mats to measure sitting on a couch or lying in bed; passive infrared sensors to detect motion in a specific area; contact switches for open–close states of doors and cupboards; and float sensors to measure the toilet being flushed. Figure 1 provides an overview of one of the employed sensors systems.

As previously stated, the system was installed in three different home settings. For each home setting, a dataset containing raw measures of the actions of the corresponding user was recorded. Data collected when more than one person were in the house were excluded. These datasets are named and described in Table 1.
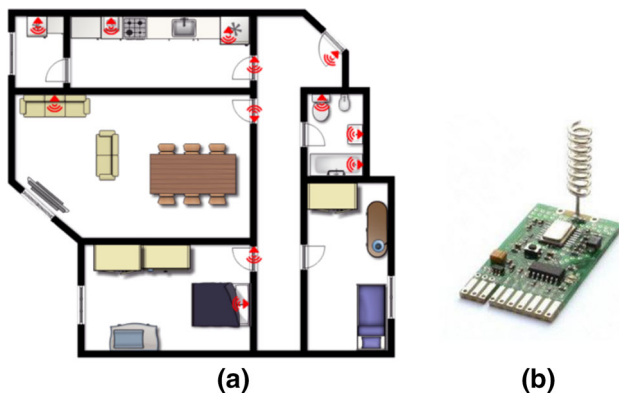
Fig. 1 Proposed monitoring system for the subject living alone: **a** floor plan of one of the home environments ('DatasetB') and locations of the sensors; **b** general view of the RFM wireless sensor nodes

Table 1 Home settings description

	DatasetA	DatasetB	DatasetC
Description	Adult man	Elderly woman	Elderly woman
Setting	House	Apartment	House
Rooms	4	2	5
Days	14 days	25 days	21 days
Sensors	12	12	12

Users' profiles were selected according to their level of autonomy in order to introduce variability in the ground truth. The different profiles of the selected candidates are as follows:

1. A completely autonomous and independent adult man ('DatasetA').
2. An elderly woman diagnosed with Parkinson's disease ('DatasetB').
3. An autonomous elderly woman with no life-threatening diseases ('DatasetC').

3 Methodology

In this study, we have defined anomalous conditions based on three different probabilities related to the sensors system. We postulate that such probabilities can represent the behavioral patterns of a person, modeling the distribution of the next features: (1) when each sensor is activated, (2) in which order, and (3) for how long.

We use these distributions as priors, meaning that they reflect our belief, a priori, of what the parameter values (and therefore the behavior of our subject) should be. Bayes rule provides the mechanism by which our prior beliefs can be converted into posterior beliefs when new

data arrives. These behavior features are modeled using Bayesian statistics because, when proper prior information is available, Bayesian estimators typically provide more accurate results with less bias and smaller mean-squared errors between the true and the estimated values than the maximum likelihood estimators. Table 2 shows the distributions employed to model our knowledge, and the corresponding prior distributions. The precise distribution of our features is always unknown (because we never have infinite data to learn that distribution), but, by marginalizing out the parameters, we can compute the expectation of the distribution instead.

3.1 Features for behavioral modeling and anomaly detection

The three features we propose in this paper to represent the behavior of a person can provide relevant information by which to judge the abnormality of human actions and behavioral patterns. The sensor activation likelihood (SAL) represents circadian rhythms and hence is of potential value for long-term health status monitoring. The sensor sequence likelihood (SSL) may be helpful in identifying confusional states or delirium. The sensor event duration likelihood (SDL) is related to the subject's physical conditions, such as weaknesses, falls, and unresponsive statuses. Besides, any of them can be very helpful to determine if the subject suffers from cognitive disorder.

We consider any of these likelihoods quite informative and useful as stand-alone, although a combination of information between these features should be also considered due to the expected improvement.

3.1.1 Sensor activation likelihood : SAL

SAL estimates the likelihood of each sensor being fired at each specific time interval. Raw data streams must be processed to provide a proper temporal format; thus, the timeline is discretized into a set of time slices. The timeslices represent the measurements of the binary sensors taken at intervals that are regularly spaced with a predetermined time granularity Δt. Sensor events are denoted as x_t^i, indicating whether sensor i is fired at least once between time t and time $t + \Delta t$, with $x_t^i \in \{0, 1\}$. Therefore, in a home environment with N state-change sensors, a binary observation vector $\mathbf{x_t} = (x_t^1, x_t^2, \ldots, x_t^{N-1}, x_t^N)^T$ is defined for each time slice (see Fig. 2).

Each sensor observation is modeled as an independent Bernoulli distribution, where ρ_t^i is the parameter of the i sensor for time interval t. The likelihood of a sensor i to be fired at time interval t is then denoted by the following:

Table 2 Overview of the distributions used to model the set of features, and the corresponding prior distribution

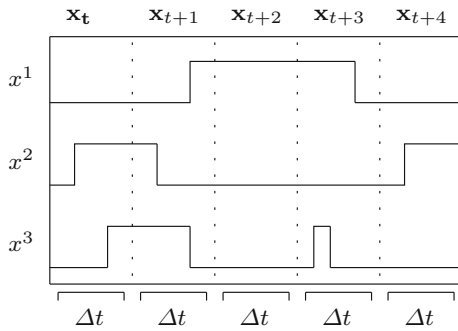| Feature	Likelihood		Prior distribution	
Name	Name	Parameters	Name	Hyperparameters
Sensor activation	Bernoulli	ρ	Beta	α, β
Sensor sequence	Multinomial	θ	Dirichlet	η
Sensor event duration	Normal	μ, λ	Normal-gamma	ϕ, κ, a, b

Fig. 2 Temporal segmentation and relation between sensor readings x^i and time intervals Δt

$$p(x_t^i|\rho_t^i) = (\rho_t^i)^{x_t^i}(1 - \rho_t^i)^{1-x_t^i} . \tag{1}$$

In this approach, we have adopted a solution for estimating the parameter ρ_t^i that takes into account not only the current interval t, but also the values of the sensor in a certain number of intervals around t. There is no point in trying to estimate the behavior of a sensor at a particular time when the activity of such sensor an instant before or after is being ignored. For this reason, we define a parameter W that provides a sort of simple sliding window method in the ρ_t^i parameter estimation process.

Since we are using a Bernoulli likelihood, the natural conjugate prior has the form of a beta distribution:

$$p(\rho_t^i|\alpha_t^i, \beta_t^i) = Beta(\rho_t^i|\alpha_t^i, \beta_t^i) = \frac{1}{B(\alpha_t^i, \beta_t^i)}\rho_t^{i\,\alpha_t^i-1}(1 - \rho_t^i)^{\beta_t^i-1} \tag{2}$$

where $B(\alpha_t^i, \beta_t^i)$ is the beta function. Hence, the posterior probability for parameter ρ_t^i is denoted by the following:

$$p(\rho_t^i|x_{\{t-W/2\}}^i\cdots x_{\{t+W/2\}}^i)p(x_{\{t-W/2\}}^i\cdots x_{\{t+W/2\}}^i|\rho_t^i)$$
$$\cdot p(\rho_t^i) \propto Beta(\rho_t^i|\alpha_t^i + R_t^i, \beta_t^i + S_t^i) \tag{3}$$

where $R_t^i = \sum_{z=t-W/2}^{t+W/2} r_z^i \xi$ and $S_t^i = \sum_{z=t-W/2}^{t+W/2} s_z^i \xi$ are respectively the effective number of observations of ones and zeros for sensor i over the window W centered at t. Being ξ a weighting factor to adjust the relevance of intervals in the window, and

$$r_t^i = \frac{1}{D}\sum_{d=1}^{D} \mathbb{1}_{\{x_t^i=1\}} \tag{4}$$

and

$$s_t^i = \frac{1}{D}\sum_{d=1}^{D} \mathbb{1}_{\{x_t^i=0\}} \tag{5}$$

the number of ones and zeros, respectively, for sensor i at interval t over the D number of days that compose the dataset; $\mathbb{1}_{\{.\}}$ denotes an indicator function. For this feature, we use the values $\alpha_0^i = \beta_0^i = 1$, resulting in a non-informative prior. Finally, by integrating out the parameters of the model, we estimate the sensor activation likelihood for sensor i, resulting:

$$\begin{aligned}\mathbf{SAL_t}(i) &= p(x_t^i = 1|x_{\{t-W/2\}}^i\cdots x_{\{t+W/2\}}^i)\\ &= \int_0^1 p(x_t^i = 1|\rho_t^i)p(\rho_t^i|x_{\{t-W/2\}}^i\cdots x_{\{t+W/2\}}^i)d\rho_t^i\\ &= \int_0^1 \rho_t^i Beta(\rho_t^i|\alpha_t^i + R_t^i, \beta_t^i + S_t^i)d\rho_t^i =\\ E[\rho_t^i] &= E[\rho_t^i] = \frac{(\alpha_t^i + R_t^i)}{(\alpha_t^i + R_t^i) + (\beta_t^i + S_t^i)}\end{aligned} \tag{6}$$

Figure 3 shows an example of the sensor activation feature over 24 h when $\Delta t = 60$ seconds and $W = 50$ (50 min).

3.1.2 Sensor sequence likelihood : SSL

The sensor sequence likelihood is estimated to predict the behavior of the user infering a plausible sensor sequence in terms of a plan structure. The sequence of sensors activated by the subject is taken as input. The first step is to extract the sequence of sensor firings, denoted by S, from the dataset. The transition probability distribution $p(s_j|s_i)$ represents the probability that the sensor s_j is fired after the activation of sensor s_i. This is given by N multinomial distributions, one for each sensor, where individual transition probabilities are denoted as $\theta_{ij} \equiv p(s_t = j|s_{t-1} = i)$.

In this case, transition parameters are assumed to be distributed as a Dirichlet, so our prior is given by the following:

$$p(\vec{\theta_i}|\vec{\eta_i}) = \text{Dir}(\vec{\theta_i}|\vec{\eta_i}) = \frac{1}{Z(\vec{\eta_i})} \cdot \theta_{i1}^{\eta_{i1}-1}\cdots\theta_{iN}^{\eta_{iN}-1} \tag{7}$$

Fig. 3 Activation likelihood estimated using frequentist statistics (*solid line*) and SAL (*dotted line*) for a motion sensor installed in a sink
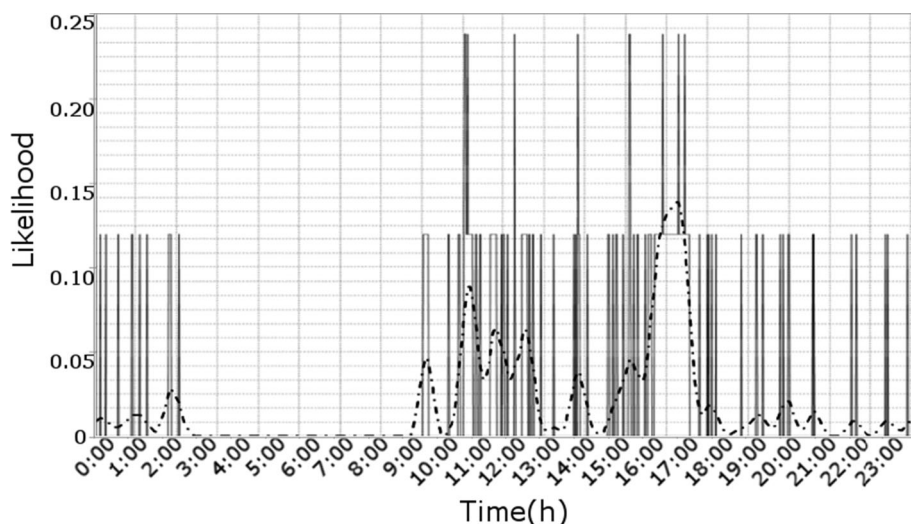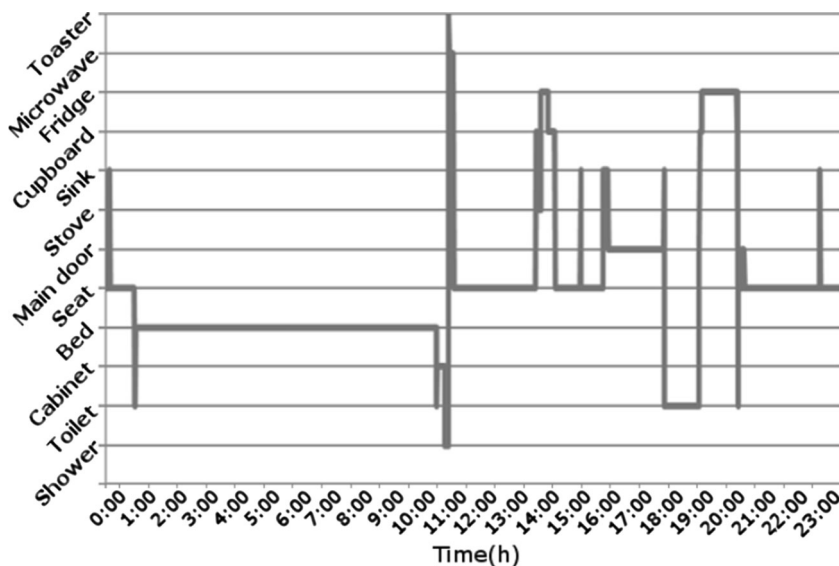

Fig. 4 Sequence S of sensor activations over a single day of data for dataset 'DatasetA'

$$Z_{\text{Dir}}(\overrightarrow{\eta_i}) = \frac{\prod_{j=1}^{N} \Gamma(\eta_{ij})}{\Gamma(\sum_{j=1}^{N} \eta_{ij})} \tag{8}$$

where the probabilities sum to one ($\sum_k \theta_{ik} = 1$). For estimating this feature, the non-informative prior is chosen as $\forall \eta_0 = 1$. Hence, the posterior probability for parameter $\overrightarrow{\theta_i}$, representing the multinomial distribution of sensor i, is denoted by the following:

$$p(\overrightarrow{\theta_i}|S, \overrightarrow{\eta_i}) = \text{Dir}(\eta_{i1} + U_{i1}, \ldots, \eta_{iN} + U_{iN}) \tag{9}$$

where $U_{ij} = \sum_{s=2}^{S} \mathbb{1}_{\{s_{t-1}=i, s_t=j\}}$ is the number of transitions from sensor i to sensor j in the firing sequence. Finally, we can determine the probability of the subject to activate the sensors in a fixed order, estimating the sensor sequence likelihood from sensor s_i to sensor s_j as follows:

$$\mathbf{SSL_i(j)} = p(\theta_{ij}|S) = \frac{\eta_{ij} + U_{ij}}{U_i + \sum_N \eta_{iN}} \tag{10}$$

Figure 4 shows an example of sensors activation sequence S over a single day of data.

3.1.3 Sensor event duration likelihood : SDL

The sensor event duration likelihood reflects the duration of the activities and movements of the subject, which is likely to be very informative for anomaly detection. First, sensor data are processed obtaining activation durations for each sensor, each duration defined in terms of a number of intervals. As previously stated, data streams are discretized into a set of time slices regularly spaced with a time granularity Δt. For each sensor i, there is a set $L_i := \{l_{i1}\ldots l_{iN}\}$ composed by N values, representing the

durations in intervals for such sensor, and $\bar{l}$ being the averaged mean of the data. The duration of each sensor is modeled as a normal distribution, where μ_i and λ_i are the mean and precision for the duration of the sensor i, respectively. Such distribution is a well-known solution for modeling human motion [20]. The likelihood is denoted by the following:

$$p(L_i|\mu_i, \lambda_i) = \frac{1}{(2\pi)^{N/2}} \lambda_i^{N/2} \exp\left(-\frac{\lambda_i}{2} \sum_{n=1}^{N} (l_{in} - \mu_i)^2\right) \tag{11}$$

In this case, we use a uniform prior with normal-gamma conjugate priors for the parameters μ_i and λ_i, so the prior of the sensor event duration likelihood is as follows:

$$NG(\mu_i, \lambda_i|\phi_i, \kappa_i, a_i, b_i) = \mathcal{N}(\mu_i|\phi_i, (\kappa_i \lambda_i)^{-1}) Ga(\lambda_i|a_i, b_i)$$
$$= \frac{1}{Z_{NG}} \lambda_i^{a_i - \frac{1}{2}} exp\left(-\frac{\lambda_i}{2}\left[\kappa_i(\mu_i - \phi_i)^2 + 2b_i\right]\right) \tag{12}$$

$$Z_{NG}(\phi_i, \kappa_i, a_i, b_i) = \frac{\Gamma(a_i)}{b_i^{a_i}} \left(\frac{2\pi}{\kappa_i}\right)^{\frac{1}{2}} \tag{13}$$

we use the values $\kappa_i = a_i = b_i > 0.1$ and $\phi_i = 0$, resulting in a non-informative prior. As in other approaches [9], we have found no differences when comparing our normal-gamma prior with other non-informative conjugate priors. Thus, the posterior probability for parameters μ_i and λ_i, representing the normal distribution for the duration of sensor i, is denoted by the following:

$$p(\mu_i, \lambda_i|L_i) = NG(\mu_i, \lambda_i|\phi_i', \kappa_i', a_i', b_i') \tag{14}$$

where

$$\phi_i' = \frac{\kappa_i \mu_i + N\bar{l}}{\kappa_i + N} \tag{15}$$

$$\kappa_i' = \kappa_i + N \tag{16}$$

$$a_i' = a_i + N/2 \tag{17}$$

$$b_i' = b_i + \frac{1}{2} \sum_{n=1}^{N} (l_n - \bar{l})^2 + \frac{\kappa_i N(\bar{l} - \mu_i)^2}{2(\kappa_i + N)} \tag{18}$$

At this point, we can compute the likelihood for sensor i of being active during l time intervals. The sensor event duration likelihood is estimated using a Student's t distribution as follows:

$$\mathbf{SDL_i(l)} = p(l|L_i) = t_{2a_i'}\left(l|\phi_i', \frac{b_i'(\kappa_i' + 1)}{a_i' \kappa_i'}\right) \tag{19}$$

3.1.4 Anomalies classification

In order to estimate whether an output of any of the presented features is anomalous or not, we define a Bayesian

credible interval (BCI). Such mechanism has been successfully employed in other anomaly detection approaches [5]. Taking the values of each likelihood at each time interval, anomalies can be detected sequentially as new outputs become available by constructing a BCI for each feature. The marginal posterior distributions of each feature can be used to construct BCIs for the set of values obtained from the training data. The p % BCI indicates that the posterior probability of the current value likelihood is falling within the interval p; thus, the BCI defines the range of plausible values for the different features. Hence, any values that fall outside of the p % BCI are classified as anomalous. The $100(1 - \alpha)\%$ BCI for a new value can be calculated as follows:

$$\bar{x} \pm z_{\alpha/2} \sqrt{\sigma^2} \tag{20}$$

where $z_{\alpha/2}$ is the $100(1 - \alpha/2)$th percentile of a normal distribution, and σ^2 is the variance of the marginal posterior distribution in the values prediction. It must be noted that in some cases (for instance, with the sensor sequence likelihood), the values considered as anomalies are only those under the range of plausible values (a sensor transition is only anomalous when it is very unlikely, not the opposite).

4 Proposed approach validation

To validate the proposed approach, two different types of datasets were employed. In order to be a reliable anomaly detection system, the proposed methodology must be verified with real data. But, due to the complexity in properly identifying all of the anomalous events that may exist in the real data, synthetic data are also included in the evaluation as proposed by other authors [5, 15], thus obtaining further validation. Hence, both real and synthetic data were employed to evaluate the performance of the method in real and possible anomalous situations. Real data obtained from the sensors were analyzed, and after being coached by an expert, the abnormal sensor readings had to be manually identified. Synthetic data were generated as a complement of the real datasets to compare the performance of the anomaly detection method developed in this study. Taking a real dataset as base, synthetic errors were generated according to the equation

$$D* = D \pm \Delta \tag{21}$$

where $D*$ is the anomalous measurement, D is the true measurement, and Δ is an offset. Using this mechanism, synthetic anomalies were randomly introduced into each dataset independently with a frequency of 5 %.

Therefore, for each one of the datasets described in Table 1, we have created a corresponding synthetic dataset where, apart from the real anomalies, a collection of anomalous measurements have been manually introduced. Such mechanism is useful to verify whether the performance of our approach increases asymptotically as more anomalous data are added.

It must be noticed that we have not generated the synthetic datasets from scratch, but they are a modified version of the datasets we have obtained using our instrumented environments. So, for each real dataset, a synthetic version of such dataset exists, where the number of anomalies is considerably higher.

4.1 Setup

Our Bayesian approach was validated splitting the original data into a test and training sets using a 'leave one day out' approach. In this approach, one full day of sensor readings is used for testing and the remaining days are used as training data. The process is then repeated for each day and the average performance measure reported. For these experiments, the data were segmented in intervals of lenght $\Delta t = 60$ seconds. In previous studies related to the area, such interval length has been proved to be long enough to be discriminative and short enough to provide good results [17].

4.2 Parameter optimization

The first experiment performs an individual analysis of the features proposed in our approach. To test the performance of the features, both real and synthetic datasets are employed. The three features are evaluated individually using three metrics: precision, recall, and F measure. The precision represents the proportion of predicted anomalies which actually correspond with an abnormal behavior, while the recall measures the proportion of actual anomalies which are correctly identified.

The use of these metrics assumes a two-class (normal and abnormal) problem and calculates the measures with respect to the number of true positives [10]. Precision and recall can be calculated using the confusion matrix shown in Table 3 that reports the number of false positives, false negatives, true positives, and true negatives. Putting it into words, the true positives represent the number of anomalies correctly identified, the false positives represent the number of normal events incorrectly classified as anomalies, the true negatives represent the number of normal events correctly rejected as anomalies, and the false negatives the number of anomalies incorrectly classified as normal events.

Precision and recall are often combined into a single measure known as the F measure, which is a weighted

Table 3 Confusion matrix in a two-class classification problem

	Inferred	
	Normal	Anomalous
Actual		
Normal	True positive (TP)	False negative (FN)
Anomalous	False positive (FP)	True negative (TN)

average of the two measures [8]. Precision and recall metrics are defined as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{22}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{23}$$

while the F measure is defined as follows:

$$F\text{measure} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{24}$$

In our approach, the parameter BCI determines the trade-off between the proportion of actual anomalies and non-anomalous patterns which are correctly identified. Tables 4, 5, and 6 show the performance values for the three different datasets ('DatasetA,' 'DatasetB,' and 'DatasetC,' respectively) when evaluated using BCIs of 80, 85, 90, 95, and 98 %. It can be noticed how, as expected, the precision increases as BCI narrows. In general terms, the performance improves when synthetic data are added to the datasets, due to the incorporation of inherently anomalous sensor activations.

The overall performance is highly influenced by the BCI. The SAL obtains the best results consistently when using 95 % BCI for both types of data. However, the performance of the SSL feature is determined by the data employed requiring a high BCI to obtain a relevant precision. The SDL, in general terms, performs better using 98 % BCI for real data and 80 % BCI for synthetic data. Such feature is quite sensitive to the distribution of the data, and when artificial anomalies (and hence easier to detect) are included, the recall is severely affected.

5 Results

The main contribution of this study is to use three features to assess the daily living pattern of the elderly and develop an abnormality detection method using Bayesian probabilistic modeling. The next experiment is aimed to evaluate the approach when the values of the three features are combined as a single output. When combining

Table 4 Individual experiments of the features for dataset 'DatasetA'

BCI	Dataset 'DatasetA'								
	SAL			SSL			SDL		
	Precision	Recall	*F* measure	Precision	Recall	*F* measure	Precision	Recall	*F* measure
Real									
98	0.92	0.59	0.72	0.34	0.62	0.44	0.86	0.99	0.93
95	0.72	0.99	0.83	0.29	0.62	0.40	0.71	0.99	0.83
90	0.37	0.99	0.54	0.29	0.68	0.41	0.60	0.99	0.75
85	0.25	0.99	0.40	0.28	0.81	0.42	0.43	0.99	0.60
80	0.19	0.99	0.32	0.26	0.87	0.40	0.39	0.99	0.56
Synthetic									
98	0.96	0.56	0.71	0.41	0.71	0.52	0.92	0.16	0.27
95	0.89	0.89	0.89	0.32	0.24	0.43	0.92	0.25	0.40
90	0.77	0.90	0.83	0.33	0.78	0.46	0.85	0.32	0.47
85	0.69	0.98	0.81	0.31	0.84	0.45	0.83	0.42	0.56
80	0.63	0.99	0.77	0.29	0.90	0.44	0.80	0.49	0.61

Table 5 Individual experiments of the features for dataset 'DatasetB'

BCI	Dataset 'DatasetB'								
	SAL			SSL			SDL		
	Precision	Recall	*F* measure	Precision	Recall	*F* measure	Precision	Recall	*F* measure
Real									
98	0.95	0.41	0.57	0.90	0.25	0.39	0.99	0.79	0.86
95	0.77	0.73	0.75	0.91	0.34	0.49	0.25	0.99	0.40
90	0.45	0.91	0.60	0.87	0.40	0.55	0.11	0.99	0.20
85	0.32	0.94	0.48	0.87	0.43	0.58	0.10	0.99	0.18
80	0.25	0.94	0.39	0.82	0.45	0.58	0.08	0.99	0.15
Synthetic									
98	0.98	0.68	0.80	0.99	0.28	0.44	0.90	0.25	0.40
95	0.93	0.86	0.90	0.90	0.43	0.59	0.90	0.34	0.49
90	0.82	0.94	0.88	0.69	0.59	0.63	0.87	0.41	0.56
85	0.69	0.98	0.81	0.31	0.84	0.45	0.83	0.42	0.56
80	0.63	0.99	0.77	0.29	0.90	0.44	0.80	0.49	0.61

every feature, the system will consider as an anomaly each time interval where the outcome of any feature is outside the range of plausible values defined by the BCI. In this experiment, for further evaluation and comparison purposes, four describing values were calculated from each dataset to analyze the performance of the abnormality detection algorithm. In this problem domain, the size of the positive data is usually small in proportion if compared with that of the negative class. In cases like this, where unbalanced data are present, it is common to use ranking-based evaluation metrics such as area under the receiver operating characteristic (ROC) [2]. Such area under the curve (AUC) is defined in terms of true-positive

rate (sensitivity) in function of the false-positive rate (specificity). Sensitivity and specificity are defined as follows:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{25}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{26}$$

It can be noticed how sensitivity is equivalent to recall. Although the accuracy is not a proper measure when dealing with unbalanced dataset, it is also included in the comparative for further evaluation. Accuracy is defined as follows:

Table 6 Individual experiments of the features for dataset 'DatasetC'

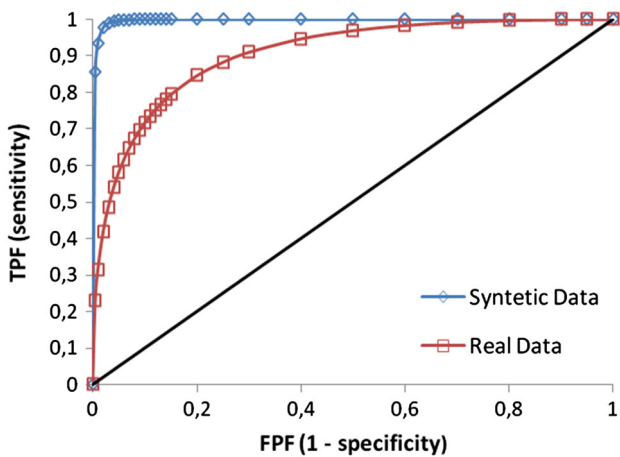| BCI	Dataset 'DatasetC'								
	SAL			SSL			SDL		
	Precision	Recall	F measure	Precision	Recall	F measure	Precision	Recall	F measure
Real									
98	0.91	0.52	0.66	0.24	0.88	0.38	0.79	0.50	0.62
95	0.87	0.83	0.85	0.92	0.32	0.48	0.26	0.66	0.37
90	0.75	0.95	0.84	0.84	0.49	0.62	0.25	0.82	0.39
85	0.67	0.97	0.79	0.73	0.56	0.64	0.27	0.90	0.42
80	0.60	0.98	0.75	0.68	0.62	0.64	0.30	0.98	0.45
Synthetic									
98	0.86	0.38	0.53	0.72	0.99	0.84	0.76	0.47	0.58
95	0.75	0.87	0.80	0.21	0.99	0.35	0.25	0.64	0.36
90	0.47	0.96	0.63	0.11	0.99	0.20	0.25	0.81	0.38
85	0.32	0.98	0.48	0.06	0.99	0.12	0.26	0.89	0.40
80	0.25	0.99	0.41	0.05	0.99	0.11	0.27	0.97	0.42

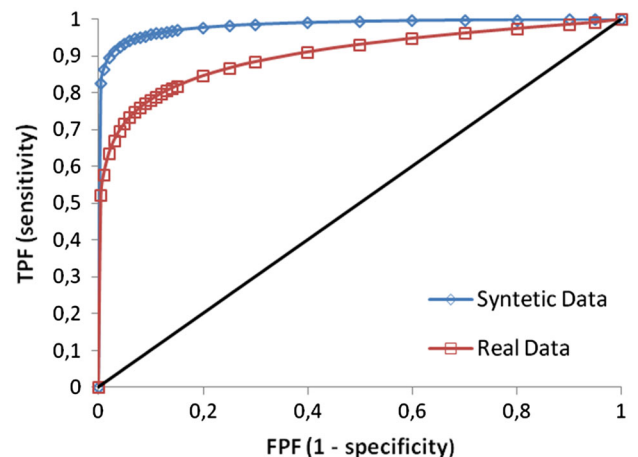**Fig. 5** ROC curve for dataset 'DatasetA,' combining the features outputs

Fig. 6 ROC curve for dataset 'DatasetB,' combining the features outputs

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (27)$$

Employing the parameters configuration obtained from the previous experiment, the performance of our Bayesian approach is then calculated. Figures 5, 6, and 7 show the results when the values of the three features are combined as a single output, for datasets 'DatasetA,' 'DatasetB,' and 'DatasetC,' respectively.

The estimated AUCs show a good performance for all datasets, and it is specially noticeable for dataset 'DatasetC.' Dataset 'DatasetA' obtains a 0.90 and 0.94 AUC for real and synthetic data, respectively. In the case of dataset 'DatasetB,' the values of the AUC are 0.90 and 0.98 for real and synthetic data, respectively. Our method obtains a

remarkable 0.99 AUC for both types of data in dataset 'DatasetC.'

It can be noticed from Table 7 how the method is quite accurate identifying negative results when using our parameter configuration. Although, in our domain, it can be more interesting to obtain a higher sensitivity to prevent anomalous situations, in that case, as shown in previous tables, by reducing the BCI, the value of the sensitivity will notably increase, but the overall performance of the method will be affected.

5.1 Feature performance example

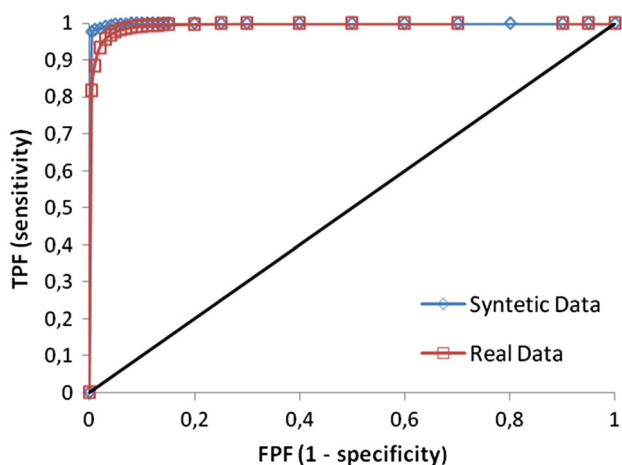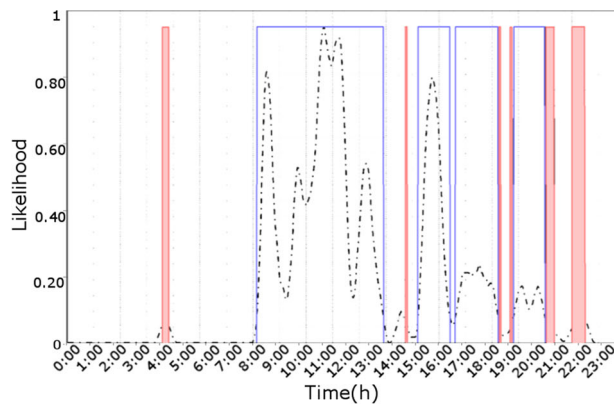In order to show how the approach can differentiate between normal and anomalous behaviors, an example

Fig. 7 ROC curve for dataset 'DatasetC,' combining the features outputs

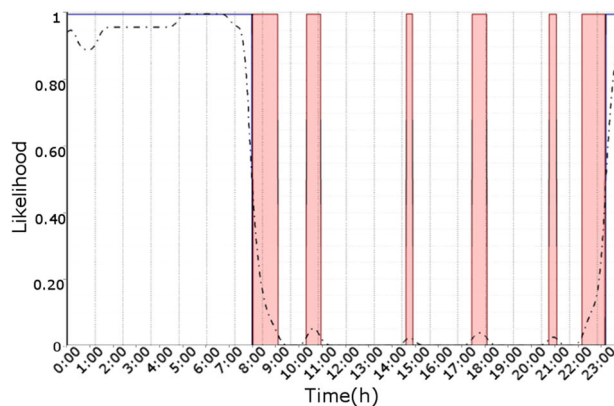Table 7 Performance of the approach for each dataset, combining the features outputs

	Specificity	Sensitivity	Accuracy	AUC
Real				
DatasetA	0.96	0.77	0.95	0.90
DatasetB	0.99	0.62	0.99	0.86
DatasetC	0.99	0.81	0.99	0.96
Avg	0.98	0.73	0.97	0.90
Synt				
DatasetA	0.99	0.83	0.98	0.96
DatasetB	0.99	0.84	0.99	0.96
DatasetC	0.99	0.85	0.99	0.97
Avg	0.99	0.84	0.98	0.96

Fig. 8 Performance examples of the SAL feature for dataset 'DatasetB' using real anomalies. Shaded periods of time represent intervals where anomalous behaviors are identified. a PIR sensor of the main door, b pressure sensor of the bed

using the SAL feature with the 'DatasetB' dataset is given in Fig. 8. We have chosen such dataset because it was generated by an elderly woman diagnosed with Parkinson's disease, and we represent the sensor activation likelihood because this feature is more illustrative and easier to plot. When people diagnosed with Parkinson's disease start to take medication, they usually notice that their symptoms go away for hours (ON times), but the symptoms return after a time (OFF times). During OFF periods, patients experience stiffness and lack of muscular coordination. During ON times, patients feel relatively fluid, clear, and in control of their movement. The fluctuations between 'ON–OFF' periods are unpredictable and represent dangerous states.

In Fig. 8b, we can see the averaged value of the SAL feature over all data for the pressure sensor installed in the bed. From 11:00 to 21:00, there are four activity peaks that strictly correspond with OFF periods of the user along different days. It can be noticed how the system identifies such periods as anomalous behaviors, because the

likelihood of using the bed at that time in a normal day is very low.

In addition, Fig. 8a shows the value of the SAL feature for the motion sensor installed in the main door. Apart from the anomalies detected close to the threshold defined by the BCI, in this case, a clear anomaly represented by the activation during the night is noticeable, between 03:00 and 4:00. Such activation matches with a time when the user had to leave the house to be hospitalized and highlights how the approach can be combined with an alarm system.

5.2 Comparative

Our last experiment was aimed to compare the performance of our model with another system that has been shown to perform well in a similar domain. In our comparative, we introduce the approach proposed by Shin et al. [15], where also three factors are calculated in order to detect anomalous behavior of elderly people living alone. These factors are activity level, mobility level, and non-

Table 8 Comparative between our Bayesian approach and the system proposed by Shin et al. [15]

Approach	Real data		Synthetic data	
Features	Specificity	Sensitivity	Specificity	Sensitivity
SAL & SDL & SSL	0.98	0.73	0.99	0.84
NRI & AL & ML	0.85	0.74	0.85	0.86

response interval. A SVDD method was used to classify normal behavior patterns and to detect anomalous behavioral patterns based on such features.

Although the datasets employed in the validation are different, the focus of both approaches is similar, and the domain and metrics employed to evaluate the systems are exactly the same. Therefore, we consider this comparative very useful to determine whether our results are satisfactory and comparable to the results obtained with other similar approaches.

Table 8 shows the results of the comparative between both systems. The first row shows the performance of our approach where SAL, SDL, and SSL are employed to model the data and the BCI to distinguish between normal and anomalous behaviors. The second row shows the performance of the approach that uses activity level, mobility level, and non-response interval as features and a SVDD to classify the behaviors. It can be noticed how the specificity is significantly better in our case, and the sensitivity is slightly better in the other approach.

A two-tail Student's t test at a confidence interval of 95 % reveal how there are no significant differences in terms of sensitivity, while differences in specificity are indeed statistically significant.

6 Discussion

The automated behavior analysis system presented here is geared to monitor the behavioral patterns of an individual living alone and does not handle the presence of multiple users.

Experimental evaluation shows not only that the proposed approach offers a very good performance identifying anomalous events, but also that in terms of specificity, it performs even better than other established methods, as the system proposed by Shin et al.

This anomaly detection method is primarily focused on detecting when something unusual is happening in the behavior of the user, but more research is still needed to establish when these anomalous behaviors are truly significant from the medical point of view. Given the nature of the Bayesian method employed, the behavior anomalies are

identified statistically and the system cannot obtain enough semantic information to provide more details about if an anomaly represents that the health of the user is whether recovering or deteriorating. However, the method allows for a simple tuning of the sensitivity and specificity that will allow for personalization of the level of service depending on the situation of the user (from healthy people to people with cognitive impariment). In an integral system for assisting elderly and disabled living alone, other information coming from the home and the healthcare system should be used to interpret the information coming from our model, and this could include holiday periods, clinical appointments, recovery periods after hospital admissions or surgery, and other information.

Nevertheless, the anomalies analysis provided by this approach is very suitable to be integrated as an intelligence provider service into a more complex telecare system, as the one presented by Williams et al. [21]. Defining several threshold levels, it is quite straightforward the adaptation of the statistical output of this approach into an alarm-based warning system. Besides, the continuous behavior supervision devised is much more sensible than current telecare services, as it detects everyday needs, not only urgent need for help, and provides a feeling for safety and emotional support to both the older adult and the family caregivers.

One limitation of the model lies in its inability to identify periodic variation of users routine behaviors, such as adopting one routine on weekdays and another on weekends. Although this problem can be addressed by separating the data appropriately, how to make this process automatic remains an open problem.

7 Conclusions

We have presented an anomaly detection method based on Bayesian statistics that can be effectively applied for identifying anomalous human behavioral patterns. The method relies on an approximate estimation of the living patterns of the user, and on prior knowledge that reflects our belief, a priori, of how such patterns should be. We have defined three probabilistic features to model the behavior of a person, whose value and efficacy as anomaly detection methods are illustrated using a case study, involving three datasets collected from different instrumented real home environments. The features presented in this work provide valuable insights about behavioral patterns of the monitored cohort, and Bayesian statistics has resulted in a very consistent way to reason under the uncertainty of human behavior.

Since we use Bayesian statistics, our approach easily integrates any known prior knowledge about the living patterns of the user, and new data arriving from the sensors

can be incorporated in near real time into the model. The whole learning process is quite fast, since we do not have to deal with convergence issues.

We concede that the amount of data to evaluate the models should be (and will be) increased; nonetheless, we emphasize that our system offers a good performance in every dataset and the features presented represent a promising approach. Moreover, the evaluation showed a consistent performance even when inherently anomalous synthetic data were introduced. The results suggest that our monitoring system and the Bayesian approach presented in this work can be a valuable framework in a home monitoring system.

References

1. Aztiria A, Farhadi G, Aghajan H (2012) User behavior shift detection in intelligent environments. In: Bravo J, Hervás R, Rodríguez M (eds) Ambient assisted living and home care. Lecture notes in computer science, vol 7657. Springer, Heidelberg, pp 90–97
2. Bradley AP (1997) The use of the area under the roc curve in the evaluation of machine learning algorithms. Pattern Recognit 30(7):1145–1159
3. Corchado JM, Bajo J, Tapia DI, Abraham A (2010) Using heterogeneous wireless sensor networks in a telemonitoring system for healthcare. Trans Inf Technol Biomed 14:234–240
4. Englebienne G (2011) Bayesian methods for the analysis of human behaviour. In: Salah AA, Gevers T (eds) Computer analysis of human behavior. Springer, London, pp 3–20
5. Hill DJ, Minsker BS, Amir E (2007) Real-time Bayesian anomaly detection for environmental sensor data. In: Proceedings of the 32nd conference of IAHR
6. Iglesias JA, Angelov P, Ledezma A, Sanchis A (2010) Human activity recognition based on evolving fuzzy systems. Int J Neural Syst 20:355–364
7. Jafari R, Encarnacao A, Zahoory A, Dabiri F, Noshadi H, Sarrafzadeh M (2005) Wireless sensor networks for health monitoring. In: Proceedings of the second annual international conference on mobile and ubiquitous systems: networking and services. MobiQuitous'05. IEEE Computer Society, Washington, DC, pp 479–781
8. Japkowicz N, Stephen S (2002) The class imbalance problem: a systematic study. Intell Data Anal 6(5):429–449
9. Joshi A, Van de Peer Y, Michoel T (2008) Analysis of a gibbs sampler method for model-based clustering of gene expression data. Bioinformatics 24(2):176–183
10. Kasteren TV, Alemdar H, Ersoy C (2011) Effective performance metrics for evaluating activity recognition methods. ARCS 2011
11. Milenkovic A, Otto C, Jovanov E (2006) Wireless sensor networks for personal health monitoring: Issues and an implementation. Comput Commun 29:2521–2533
12. Palaniappan A, Bhargavi R, Vaidehi V (2012) Abnormal human activity recognition using svm based approach. In: 2012 International conference on Recent trends In information technology (ICRTIT). IEEE, pp 97–102
13. Patterson DJ, Liao L, Fox D, Kautz H (2003) Inferring high-level behavior from low-level sensors. In: Dey AK, Schmidt A, McCarthy JF (eds) UbiComp 2003: Ubiquitous computing. Lecture notes in computer science, vol 2864. Springer, Heidelberg, pp 73–89
14. Pynoos J (2002) Neglected areas in gerontology: housing adaptation
15. Shin JH, Lee B, Park KS (2011) Detection of abnormal living patterns for elderly living alone using support vector data description. IEEE Trans Inf Technol Biomed 15(3):438–448. doi:10.1109/TITB.2011.2113352
16. Suzuki R, Ogawa M, Otake S, Izutsu T, Tobimatsu Y, Iwaya T, Izumi SI (2006) Rhythm of daily living and detection of atypical days for elderly people living alone as determined with a monitoring system. J Telemed Telecare 12(4):208–214
17. van Kasteren TLM, Noulas A, Englebienne G, Kröse BJ (2008) Accurate activity recognition in a home setting. In: Proceedings of the 10th international conference on Ubiquitous computing, UbiComp '08ACM, New York, NY, USA, pp 1–9
18. Virone G, Noury N, Demongeot J (2002) A system for automatic measurement of circadian activity deviations in telemedicine. IEEE Trans Biomed Eng 49(12):1463–1469
19. Virone G, Alwan M, Dalal S, Kell SW, Turner B, Stankovic JA, Felder RA (2008) Behavioral patterns of older adults in assisted living. IEEE Trans Inf Technol Biomed 12(3):387–398
20. Wang JM, Fleet DJ, Hertzmann A (2008) Gaussian process dynamical models for human motion. IEEE Trans Pattern Anal Mach Intell 30(2):283–298
21. Williams G, Doughty K, Bradley DA (1998) A systems approach to achieving carernet-an integrated and intelligent telecare system. Trans Inf Technol Biomed 2(1):1–9. doi:10.1109/4233.678527
22. WorldHealthOrganization (2007) Global age-friendly cities: a guide. World Health Organization
23. Wu X, Ou Y, Qian H, Xu Y (2005) A detection system for human abnormal behavior. In: IEEE/RSJ international conference on untelligent robots and systems, 2005 (IROS 2005). IEEE, pp 1204–1208
24. Yin J, Yang Q, Pan JJ (2008) Sensor-based abnormal human-activity detection. IEEE Trans Knowl Data Eng 20(8):1082–1090
25. Zola IK (1997) Living at home: the convergence of aging and disability. Baywood Publishing, Amityville, pp 25–39