

# An Ontology and Pattern Clustering Approach for Activity Recognition in Smart Environments

K. S. Gayathri, Susan Elias and S. Shivashankar

**Abstract** Activity recognition aims at modeling the occupants' behavior by analyzing the sensor data collected from the smart environment. Though most of the activity recognition systems use supervised learning techniques for building such models there is a shift towards the unsupervised learning paradigm as the process of annotating and labeling the data is prone to errors. This paper proposes an *Event Pattern Activity Modeling Framework (EPAM)* to identify the occupant activity pattern from the sensor data by using an unsupervised machine learning approach and further analysis is done with a knowledge driven approach. In the context of smart environments, an activity is considered as a sequence of events that are generated continuously from the sensor data. The segmentation algorithm proposed in *EPAM* is used to identify appropriate event patterns for an activity that are then grouped together using a pattern clustering algorithm that presents a hierarchy of activities. The set of activities of the occupant, observed in a smart environment is not always sequential but is highly interleaved and discontinuous. The proposed algorithm accommodates this valid factor by an innovative use of the *Jaro Winkler* similarity measure. The hierarchy of activity generated by the pattern clustering approach is used for activity modeling. Ontology based activity modeling is preferred over other modeling techniques because of its unified modeling, representation and semantically clear reasoning. The experimental

---

K. S. Gayathri (✉) · S. Elias  
Department of Computer Science and Engineering,  
Sri Venkateswara College of Engineering,  
Sriperumbudur 602117, India  
e-mail: [gayasuku@svce.ac.in](mailto:gayasuku@svce.ac.in)

S. Elias  
e-mail: [susanelias70@gmail.com](mailto:susanelias70@gmail.com)

S. Shivashankar  
Ericsson Research India, Chennai, India  
e-mail: [s.shivashankar@ericsson.com](mailto:s.shivashankar@ericsson.com)

results show that the proposed *EPAM* framework of segmentation, pattern clustering and ontological modeling is efficient and more effective than the existing approach of activity modeling.

## 1 Introduction and Related Work

The concept of smart environment has evolved over the years into applications that are very challenging to design, develop and deploy [1–3]. The availability of low cost sensors, paradigms such as the Internet of Things (IoT), and research contributions from the fields of machine learning, ambient intelligence and ubiquitous computing together help in the conceptualization of smart environments [4]. The key to smart environments lies in capturing the data in real time from the environment and in applying reasoning dynamically to achieve a specified goal. Smart environment applications can be integrated in socially relevant scenarios that would have a profound impact on the future of society. Some of the popular smart home applications range from elder and child care [5–7] energy efficiency, surveillance, offices, hospitals, agriculture, disaster assessment, to name a few. The focus of our research work is in the building of a reasoning system for smart environments. The application use case in this research work is the design of a smart home [8] that assists the occupant with *Activities of Daily Living (ADL)*. Activity recognition is the key part in a reasoning system and it is built by training the system on occupants' behavior. The trained system or the activity model can then be used for detection, prediction and decision making [1].

Learning the behavioral pattern of the occupant is essential for effectively reasoning and training the activity recognition system. Information on the normal activities can be obtained from various sources such as data from previous observations or from domain experts, text corpus and web services in specific cases [9–11]. Besides video and audio capturing devices, sensors are largely used to capture the behavior of the occupants in smart environments. Wearable and object based sensors can be efficiently utilized to continuously report about the environment in a pre-specified manner. Extensive research work is being carried out in video based activity recognition, but it has the limitation of violating the privacy of the occupants' especially in smart home applications [9]. Hence the current research focus is towards sensor based activity recognition which includes the generation of models from the data collected from these sensors. Sensor based activity recognition system modeling can be carried out using two approaches namely Data driven and Knowledge driven approaches and this classification is based on the source of the information and the strategies used for training [1, 9]. Data driven approaches analyze the data that describes the occupants' behavior collected from the various sensors deployed in the environment [3, 10]. Machine learning approaches are then used to build an activity model that further learns from the occupants' behavioral data. Sensor data used for training the activity model will either be annotated or unlabelled. Supervised learning techniques [3, 9]

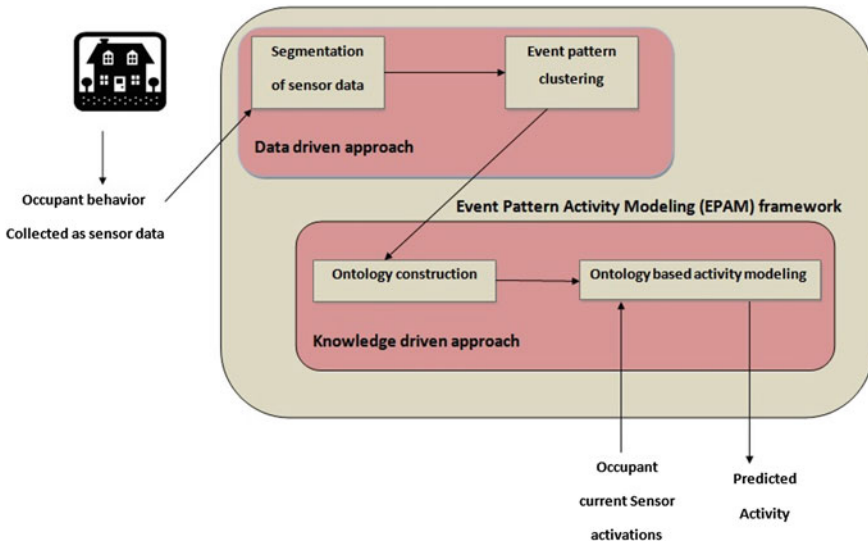
are applied if the data is labeled and unsupervised or semi supervised techniques [12, 13] are applied to unlabelled data. Semantically clear activity modeling can be achieved using knowledge driven approaches that use rich domain knowledge for activity modeling and pattern recognition [9, 14, 15]. Representation of activities in the form of knowledge helps in reusability and scalability as most Activity Daily Living (ADL) activities are similar functions for all occupants. A hybrid Data driven and Knowledge driven approach is used to build an Activity recognition model. The proposed framework employs a data driven approach for extracting the activity patterns from sensor data through behavioral clustering and variability analysis and then models the extracted activity pattern through an ontology based Knowledge driven approach.

## 2 Proposed Work

The information on *ADL* plays a vital role in building an activity recognition system for assisted living. This demands the mining of sensor data for various activities of an occupant using data mining and machine learning techniques. The proposed Event Pattern Activity Modeling (EPAM) framework is shown in Fig. 1 and it works on an unsupervised machine learning technique to extract meaningful activity patterns from large volumes of unlabelled sensor data and performs activity modeling using an ontology based technique.

### 2.1 Segmenting Sensor Data

Sensors deployed in smart environments record information about both the occupant and the environment. The received sensor data is in the form of events that are represented as a set containing information such as date, time, sensor identification and status. Event pattern represents an activity and are mined from sensor data through appropriate segmentation approach. Segmentation of sensor data can be done using various approaches: Fixed time interval based, Variable time interval based, Dominant sensor based and Location based. Fixed interval based approach used in [8], segments the sensor data at fixed time interval for generating event patterns. This approach has the drawback of omitting some useful information for an activity if the activity prolongs for more than the fixed time period or it may combine two activity patterns together if the fixed time interval is large than the activity duration. Hence a fixed time interval based approach is not appropriate for segmenting sensor information in smart home applications. Variable interval based approach [14] can be used to vary the time period based on the activity. Since the sensor data is unlabelled and the domain knowledge related to the duration of activity is not available, the time period cannot be varied for different situations. Dominant sensors [16] are sensors that is very important for



**Fig. 1** Event pattern activity modelling (EPAM) framework

each activity. E.g. In “sleep activity”, bed sensors should be ‘ON’ though there are other co existing sensor activations and therefore bed sensors are dominant for sleeping activity. Event patterns are generated in such a way that they contain sequence of events collected with at least one dominant sensor. It is not always possible to define dominant sensors for all activities since an unsupervised approach is used. Another approach is location based where a majority of occupant activity is location specific such as sleeping is associated with bedroom, cooking is done in kitchen and so on. The sensor sequences are segmented based on their spatial location and it will result in event patterns that will be more appropriate for a smart home scenario. The proposed Algorithm 1—Segmentation algorithm takes the continuous sensor data stream as the input and outputs the various event patterns generated based on the spatial location. The line 6 in Algorithm 1 first makes a comparison as to whether any two subsequent events occur in the same location. In such a case, using line 9 in the algorithm they are appended to the sequence of events pertaining to that location else they get represented as event sequences belonging to other relevant spatial locations. This segmentation helps in identifying all activities within a specific location. The segmented sequences are maintained on different data structures for different location.

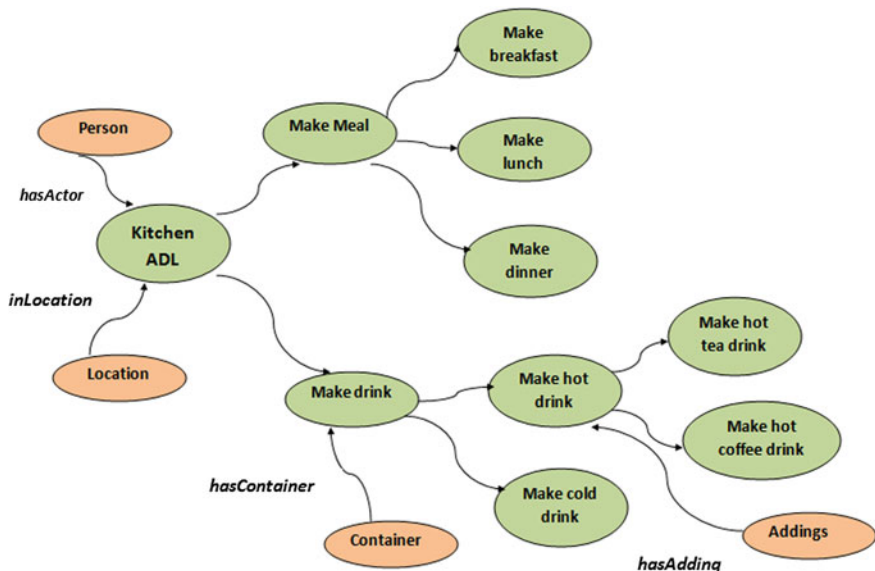


Fig. 2 Hierarchy of kitchen ADL activities

---

**Algorithm 1:** Segmenting sensor data

---

```

1 Procedure Segmenting Sensor data(Sensordata)
2 Sensordata is a sequence of event spatiotemporaldata
3 Initialise the spatial tables to empty
4 Initialise sp.spatio as empty for all spatial location
5 foreach event  $e_i$  in Sensor data do
6   if  $e_i.spatio == e_{i+1}.spatio$  then
7     | sp.spatio  $\leftarrow$  sp +  $e_i$ 
8   end
9   add sp.spatio to spatio table
10  set sp.spatio to empty
11 end

```

---

## 2.2 Event Pattern Clustering

Event pattern clustering algorithm handles unlabelled sensor data and groups the various sequences together to identify the relevant activity of the occupant in a particular spatial region. The proposed approach identifies all the sub activities pertaining to a particular activity of the occupant and in a way follows hierarchical clustering. The reason behind preferring hierarchical clustering compared to the other clustering approach is that, the activity of the occupant in smart home can be viewed as hierarchy of activities as shown in Fig. 2. What makes it different from

conventional hierarchical technique is that it clusters similar sequences of events rather than grouping discrete events.

The following are the steps involved in event pattern clustering algorithm:

### Step 1—Initializing the clusters

The event patterns generated by the location based segmentation algorithm are taken as the input to the event pattern clustering algorithm. Each event pattern available in the spatial table is initialized as a separate cluster with event pattern as the center for the cluster.

### Step 2—Extracting temporal information from event pattern

The temporal information from event pattern is extracted by storing the start time of every event pattern with the duration of the activity. This is very essential in order to build the personal occupant profile containing the information regarding the time range around which the occupant does the activity and the duration taken for that activity.

### Step 3—Calculate the similarity measure

Similarity between event patterns can be measured by a number of similarity measures like Euclidean distance, Manhattan distance and Levenshtein edit distance [12] but all these do not address issues related to event sequences in a smart environment. In such scenarios the event patterns need not be of the same length, the events may not occur in the same order every time or there may be interleaving of event sequences between different activities. In order to handle the above mentioned issues effectively, our approach uses *Jaro Winkler* similarity measure as given in Eq. 2. The approach allows a range within which the event sequences may be discontinuous as given in Eq. 3 and no two event sequences can represent an activity if they are completely discontinuous.

$$d_j = \begin{cases} 0, & \text{if } m \text{ is } 0 \\ \frac{1}{3} \left( \frac{m}{|S_1|} + \frac{m}{|S_2|} + \frac{m-t}{m} \right) & \text{otherwise} \end{cases} \quad (1)$$

$$d_w = d_j + (l * p(1 - d_j)) \quad (2)$$

$$w_l = \frac{\max(|S_1|, |S_2|)}{2} - 1. \quad (3)$$

The distance measure is given in Eq. 1 and it takes any value between [0, 1]. It takes a maximum value of 1 if the events are completely equal, values closer to 1 represent more similar sequences and it is zero if they are unequal. Equation 2 gives the Jaro distance for strings  $S_1$  and  $S_2$ ,  $l$  is the length of common prefix at the start of a string and can take a maximum of 4 characters,  $p$  is a scaling factor that indicates how much the score is adjusted upwards for strings with common prefixes,  $m$  is the number of matching characters,  $t$  is half the number of transpositions  $w_l$  is the window limit within which the discontinuity is allowed. To understand how the *Jaro distance* measure works, consider an example as shown in Fig. 3



Fig. 3 Illustration for event sequence

where the each event is represented as a character. Two patterns are observed in the sequence of events for which  $m$  takes a value of 3,  $w_l$  is 1,  $t$  is 0, length of  $S_1$  and  $S_2$  is 5. Jaro distance is calculated as 0.6 and Jaro Winkler distance is 0.64. The result gives a similarity measure of 64 %. Using the above mentioned similarity measure a similarity matrix is constructed with the similarity value between every cluster.

**Step 4—Merging of clusters**

Clusters are to be merged to form hierarchy of activity patterns and clusters which have minimum similarity distance are merged together. After the clusters are merged the similarity measures are to be recalculated between newly formed cluster patterns and every other cluster. Lines 9–13 in the Algorithm 2—Event pattern clustering does the process of merging and recalculating similarity measure. This calculation is based on single linkage clustering approach where the link between two patterns is made by a single event pattern that is formed from two event patterns (one in each cluster) that are closest to each other. The steps 2–4 needs to be repeated until all clusters are merged into a single cluster. Since, the event patterns with respect to the spatial location is taken, the event pattern clustering algorithm is able to identify the various activities done by the occupant in that particular spatial location.

---

**Algorithm 2:** Event pattern clustering

---

```

1 Procedure Event pattern clustering(Event patterns)
2 Event patterns set of event patterns spatioeventable
3 begin
4   foreach spatial location  $s_i$  do
5     foreach event pattern  $e_i$  in the Spatial location table  $s_i$  do
6       | calculate similarity measure with other event patterns
7     end
8     set each event patterns as cluster pattern center set clusnum = no of event patterns
9     repeat
10    | merge two closest event patterns
11    | update the similarity measure for every event pattern
12    | reduce clusnum by one
13    until clusnum == 1
14   end
15 end

```

---

### 2.3 Activity Modeling: Ontology Construction

Activity modeling is required to increase the comfort of the occupant through automation which assists the occupant in performing ADL activities. There are several ways by which modeling can be done, the most commonly used approaches are Artificial Neural Network (ANN), Hidden Markov Model (HMM) and Dynamic Bayesian Network (DBN). The activity models built by the above methods cannot be reused for activity modeling for different occupants. Information about the occupant activity is usually obtained from the domain expert, which makes the system highly dependent on the domain expert. The information given by the expert may not provide the complete knowledge about the occupant activities posing a limitation to the knowledge driven approach. So, we put forth a method of extracting information about the occupant activities through a data driven approach. The extracted information is then represented using knowledge representation techniques such as logics, ontologies. The Fig. 2 shows that the activity of the occupant is hierarchical in nature, so the best approach to represent and reflect this information is through ontology representation. Ontology is preferred because of its systematic representation and its advantage of interoperability, reusability and scalability. Two levels of ontology are constructed 1. ADL ontology is constructed by representing the activities of the occupant and 2. User profile ontology is constructed to represent personal information of the occupant. The information extracted in step 2 of the event pattern clustering algorithm is used for constructing user profile ontology and clustered patterns for constructing ADL ontology. Subsumption, equivalence reasoning techniques can be applied on ontology to perform semantic reasoning for assisting the occupant in his activities.

## 3 Experimental Analysis

Sensor data used for the experimental analysis were collected from WSU CASAS smart home project [12]. Java API is used to implement the Location based segmentation and event pattern clustering algorithm. The Fig. 4 shows the event patterns that are generated by applying the location based segmentation algorithm on the smart home sensor data. The ontology is constructed using Protege, an ontology editor and FACT ++ is used for ontology reasoning. Figure 5 shows the class constructed using ontology. The efficiency of the proposed system is measured through *misclassification rate* which is the ratio between total number of incorrect classification of activity to total number of activity classification during activity recognition. The experiment was conducted with smart home data set [12]



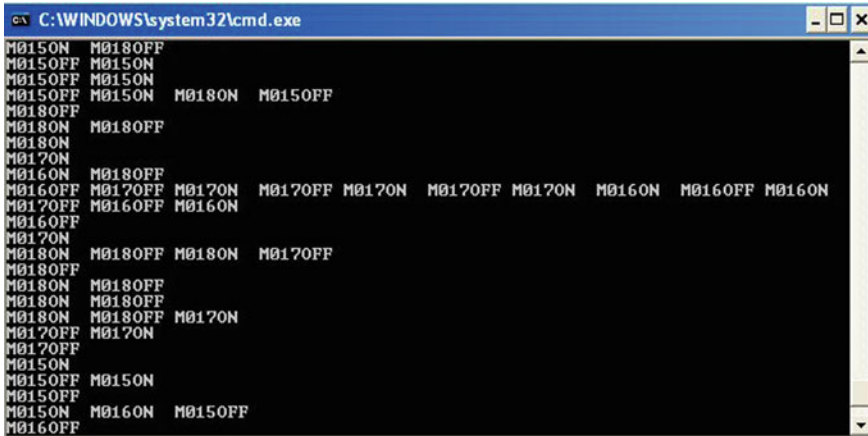


Fig. 4 Event patterns from location based segmentation of sensor data

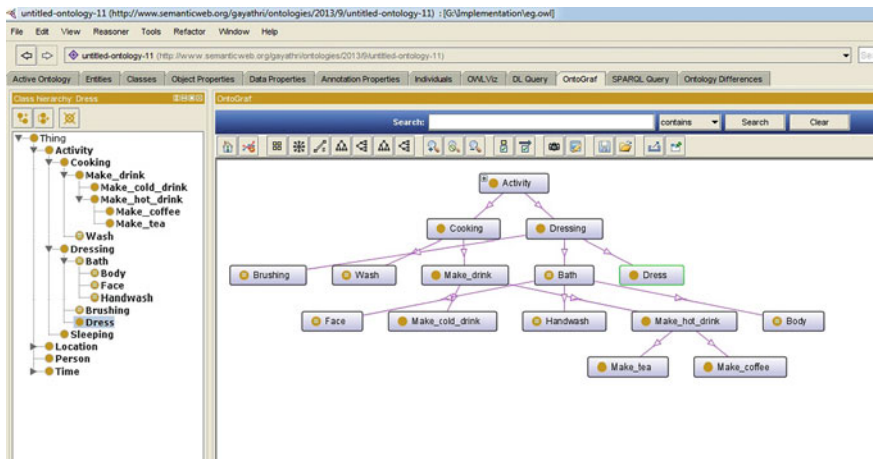
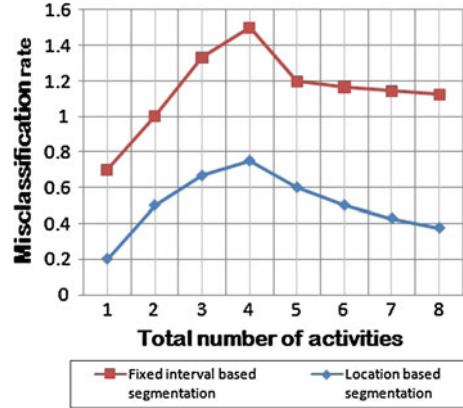


Fig. 5 Ontology construction using protege tool

and the results showed that the misclassification rate is much less in location based rather than fixed interval based segmentation as shown in Fig. 6 thus confirming location based segmentation most suitable for efficient activity modeling.

**Fig. 6** Comparison between fixed interval based and location based segmentation



## 4 Conclusion

Today sensors technology has augmented the field of ubiquitous computing with massive reduction in cost and size thereby enabling the deployment of very challenging applications. A smart environment is an application that effectively utilizes sensors to create Ambient Intelligence for automation in a wide range of domains from entertainment to health care. The most important task in such environments is the identification of occupant activity. Our proposed Event Pattern Activity Modeling (EPAM) follows unsupervised machine learning approach to identify the patterns in the unlabeled collection of sensor data through pattern clustering. We have identified that Jaro Winkler similarity measure compares patterns efficiently in smart environments. The proposed pattern clustering also organizes the clusters into a hierarchy thus identifying the various sub activities of the occupant. Ontology is constructed using the event patterns for each activity which is used for semantic reasoning of new sensor data (test input). The experimental study shows that the proposed approach of combining data driven approach and knowledge driven approach gives better prediction than using a data driven approach. The future enhancement of this work will focus on context based pattern clustering that would augment the Ambient Intelligence (AmI) in smart environments.

## References

1. Cook, D.J., Augusto, J.C., Jakkula, V.R.: Ambient intelligence: technologies, applications, and opportunities. *Pervasive Mob. Comput.* **5**(4), 277–298 (2009)
2. Cook, D.J., Das, S.K.: How smart are our environments? An updated look at the state of the art. *Pervasive Mob. Comput.* **3**(2), 53–73 (2007)

3. Augusto, J.C., Nakashima, H., Aghajan, H.: Ambient intelligence and smart environments: a state of the art. In: *Handbook of Ambient Intelligence and Smart Environments*, pp. 3–31 (2010)
4. Alam, M.R., Reaz, M.B.I., Ali, M.A.M.: A review of smart homes - past, present, and future. *IEEE Trans. Syst. Man Cybern. Part C* **42**(6), 1190–1203 (2012)
5. Lotfi, A., Langensiepen, C.S., Mahmoud, S.M., Akhlaghinia, M.J.: Smart homes for the elderly dementia sufferers: identification and prediction of abnormal behaviour. *J. Ambient Intell. Humanized Comput.* **3**(3), 205–218 (2012)
6. Corts, U., Urdiales, C., Annicchiaric, R.: *Intelligent Healthcare Managing: An Assistive Technology Approach*, pp. 1045–1051. Springer-Verlag Berlin Heidelberg (2007)
7. Augusto, J.C., Huch, M., Kameas, A., Maitland, J.: *Handbook of ambient assisted living-technology for healthcare, rehabilitation and well-being. Handbook of Ambient Assisted Living*, vol. 11. IOS Press, US (2012)
8. Sukanya, P., Gayathri, K.S.: An unsupervised pattern clustering approach for identifying abnormal user behaviors in smart homes. *Int. J. Comput. Sci. Netw.* **2**(2), 115–122 (2013)
9. Chen, L., Hoey, J., Nugent, C.D., Cook, D.J., Zhiwen, Y.: Sensor-based activity recognition. *IEEE Trans. Syst. Man Cybern. Part C* **42**(6), 790–808 (2012)
10. Aztiria, A., Izaguirre, A., Augusto, J.C.: Learning patterns in ambient intelligence environments: a survey. *Artif. Intell. Rev.* **34**(1), 35–51. Springer, Netherlands (2010)
11. Atallah, L., Yang, G.-Z.: The use of pervasive sensing for behaviour profiling—a survey. *Pervasive Mob. Comput.* **5**(5), 447–464 (2009)
12. Rashidi, P., Cook, D.J., Holder, L.B., Schmitter-Edgecombe, M.: Discovering activities to recognize and track in a smart environment. *IEEE Trans. Knowl. Data Eng.* **23**(4), 527–539 (2011)
13. Rashidi, P., Cook, D.J.: Keeping the resident in the loop: Adapting the smart home to the user. *IEEE Trans. Syst. Man Cybern. Part A* **39**(5), 949–959 (2009)
14. Chen, L., Nugent, C.D., Wang, H.: A knowledge-driven approach to activity recognition in smart homes. *IEEE Trans. Knowl. Data Eng.* **24**(6), 961–974 (2012)
15. Riboni, D., Bettini, C.: Cosar: hybrid reasoning for context-aware activity recognition. *Pers. Ubiquit. Comput.* **15**(3), 271–289 (2011)
16. Aztiria, A., Augusto, J.C., Basagoiti, R., Izaguirre, A., Cook, D.J.: Discovering frequent user–environment interactions in intelligent environments. *Personal Ubiquitous Comput.* **16**(1), 91–103 (2012)