International Journal of Modern Engineering Research (IJMER)
www.ijmer.com Vol.2, Issue.1, Jan-Feb 2012 pp-353-358 ISSN: 2249-6645

Automatic Discovery and Recognition of Activities in Smart Environment

L.Karthika*, A.C.Sumathi**

*II M.E CSE, Sri Shakthi Institute Of Engineering and Technology, Anna University, Coimbatore.

** Assistant Professor, CSE, Sri Shakthi Institute Of Engineering and Technology, Anna University, Coimbatore.

***ABSTRACT*:**

The machine learning and pervasive sensing technologies found in smart homes offers opportunities for providing health monitoring and assistance to individuals experiencing difficulties living independently at home. To monitor the functional health of smart home residents, there is a need to design technologies that recognize and track activities that people normally perform as part of their daily routines. The existing approaches are applied to activities that have been preselected and for which labeled training data are available. For this, an automated approach is proposed for activity tracking, which identifies frequent activities that naturally occur in an individual's routines. With this capability, the occurrences of regular activities are monitored and also can detect the changes in an individual's patterns.

***Keywords*—Activity recognition, Data mining, sequence mining, clustering, smart homes.**

INTRODUCTION

A convergence of technologies in machine learning and pervasive computing as well as the increased accessibility o robust sensors and actuators has caused interest in the development of smart environments to emerge. Smart Environments can assist with valuable functions such as remote health monitoring and intervention. The need for the development of such technologies is underscored by the aging of the population, the cost of formal health care, and the importance that individuals place on remaining independent in their own homes. To function independently at home, individuals need to complete Activities of Daily Living (ADLs) [1] such as eating, dressing, cooking and drinking. Automating the recognition of activities is an important step toward monitoring the functional health of a smart home resident.

In response to this recognition need, several approaches to model and recognize are designed. The generally accepted approach is to model and recognize those activities that are frequently used to measure the functional health of an individual [2]. However there is a number of difficulties in this approach. First, there is an assumption that each individual performs most, or all, standard ADLs in a consistent predefined manner in their home environments where they can be monitored. But this is certainly not always the case. In addition, the same individual might perform even the same activity indifferent ways, requiring methods that can also deal with intrasubject variability.

Second, tracking only preselected activities ignores the important insights that other activities can provide on the functional health of an individuals. This highlights the fact that it is important for a caregiver to recognize and monitor all activities that an individual regularly performs in their daily environments.

Third, to track a predefined list of activities, a significant amount of training data must be labeled and made available to the machine learning algorithm. Unfortunately, collecting and labeling such sensor data collected in a smart environment is an extremely time-consuming task.

In this paper, an unsupervised method is introduced for discovering and tracking activities in a smart environment that addresses these issues. This project based on the context of the CASAS Smart Home Project [9] by using the sensor data that are collected in the CASAS smart apartment testbed. The unsupervised nature of this model provides a more automated approach for activity recognition than offered by previous approaches.

Compared to traditional methods, for activity recognition which uses HMM model, our first approach is that discover the frequent activities by using the unique mining method, along with a clustering step to group discovered patterns. For the recognition step, an extended version of a hidden Markov model (HMM) is created to represent the activities and their variations, and to recognize those activities when they occur in the smart environment.

In the remainder of this paper, the approaches of this project are activity discovery, recognition, and tracking are explained. In section II, these approaches are compared with the related work. In section III, discovery of activities using mining and clustering methods. Then this paper's main contribution in section IV describes how discovered activities can be recognized using the HMM model.

RELATED WORK

Activity recognition needs different approaches. These approaches differ accordingly to the type of sensor data, activity model and the methods to annotate the sample data.

A. Sensor Data

The different types of sensor information are effective for classifying different types of activities. Previously [3], [4] are collected the sensor information from the state-change sensors and RFID tags. Some researchers such as [5] processed the video to recognize the activities. While accessing the video is very complex.

B. Activity Models

The numbers of machine learning models are used for activity recognition. Naïve Bayes classifiers have been used with promising results for activity recognition [6]. In this approach, Hidden Morkov Model is employed to recognize the related activities from stream of sensor information.

C. Annotation Methods

An aspect of activity recognition, a method is used to annotate the sample data. Most of the researchers have been used the labeled information for activity recognition [4], [5]. This is not always the case.

Here a new mining method is introduced called, Discontinuous Varied-Order Sequential Mining (DVSM), which is able to find frequent pattern of activities that may be discontinuous and might have variability in the ordering. Intrasubject variability issue is addressed. Activity clustering is employed here to group the patterns. In the next step the Hidden Morkov Model is used to represent the activities and their variations and to recognize those activities. The architecture of the system is shown in Fig. 1.

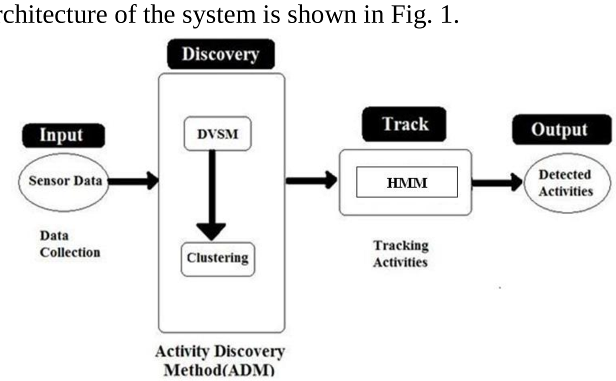


Figure.1 Main Components for discovering and tracking activities

DISCOVERING ACTIVITIES

The first step, the frequent and repeatable sequences of sensor events are considered, that comprise our smart environment's notion of an activity. By applying frequent sequential pattern mining techniques, with this contiguous events are identified, consistent sensor event sequences that might indicate an activity of interest. Ruotsalainen and Ala-Kleemola [7] introduce their Gais algorithm for detecting interleaved patterns using genetic algorithms, but this is a supervised learning approach that looks for matches to specific pattern templates. Given that sequential patterns are discoverd that may be discontinuous and have variability in the ordering, another possible approach is to cluster the sensor events.

The limitation of clustering algorithms for our problem is that all of the data points are don not clustered, but only those that are part of an activity sequence which is likely to occur frequently and with some degree of regularity or recognisability. So sequence mining and clustering algorithm is combined into an Activity Discovery Method (ADM) to identify frequent activities and cluster similar patterns together.

A. Discovering Frequent Discontinuous Sequences

DVSM is used to find discontinuous instances. This approach is different from frequent item set mining because the orders of items are considered as they occur in the data.

To find the Discontinuous Sequences from the input data D, DVSM first creates a reduced data set D_r containing the top most frequent events. Next, DVSM slides a window of size 2 across D_r to find patterns of length 2. DVSM extends the patterns discovered in the previous iteration by their prefix and suffix events, and will match the extended pattern against the already discovered patterns to see if it is a variation of a previous pattern, or if it is a new pattern.

Levenshtein (edit) distance [8] is used to define a similarity measure sim(A,B) between the two patterns. The edit distance, e(A,B), is the number of edits. The similarity measure is defined based on the edit distance as

$$sim(A,B) = 1 - \left(\frac{e(A,B)}{\max(|A|,|B|)}\right)$$

The continuity between component events, Γ_e, is defined or each two consecutive events in an instance. Then $\Gamma_e(e')$, the event continuity for e' is defined as

$$\Gamma_e(e') = \frac{1}{s_{e'}+1}$$

The instance continuity Γ_i reflects how continuous its component events are. Such that $\Gamma_i(a_i^j)$, for an instance j of a variation a_i will be defined as in, where $|a_i^j|$ is the length of a_i^j

$$\Gamma_i(a_i^j) = \frac{1}{|a_i^j|}\sum_{k=1}^{a_i^j} \Gamma_e(k)$$

The continuity of a variation, Γ_v, is then defined as the average continuity of its instances. $\Gamma_v(a_i)$ is defined as in, where n_{a_i} shows the total number of instances for variation a_i

$$\Gamma_v(a_i) = \frac{1}{n_{a_i}}\sum_{j=1}^{n_{a_i}} \Gamma_i(a_i^j)$$

The continuity,Γ_g, of a general pattern g is defined as the weighted average continuity of its variations. Γ_g is defined according to, where the continuity for each a_i is weighted by its frequency f_{a_i} and n_a shows the total number of variations for general pattern a

$$\Gamma_g(a_i) = \frac{\sum_{i=1}^{n_a} \Gamma_v(a_i) * f_{a_i}}{\sum_{i=1}^{n_a} f_{a_i}}$$

Building on this definition of continuity, patterns that are interesting and the variation of those patterns are found. The rest of the patterns and variations are pruned.

B. Clustering Sequences into Groups of Activities

The second step of the ADM algorithm is to identify pattern clusters that will represent the set of discovered activities. Specifically, ADM groups the set of discovered patterns P into a set of clusters A. The resulting set of clusters centroids represents the activities that to be modelled, recognize, and track. Though ADM uses a standard k-means clustering method [10], there is a need to define a method for determining cluster centroids and for comparing activities in order to form clusters. Two methods that are commonly used for comparing the similarity of sequences are edit distance and longest common subsequence (LCS) for simple sequences.

The patterns discovered by DVSM were composed of sensor events. In the clustering algorithm, the pattern is composed of states. States correspond to the pattern's events, but are enhanced to include additional information such as the type and duration of the sensor events. In addition, several states together to form a new state. Then all consecutive states are combined that are corresponding to the sensors of the same type to form an extended state. To calculate the similarity between two activities X and Y, compute the distance between their extended state lists and using our general edit distance to account for the state information and the order mapping frequencies. The general edit distance for two patterns X and Y can be defined based on the traditional edit distance, the order distance, and the attribute distance

The general edit distance gives us a measure to compare activities and also to define cluster centroids. Each cluster representative represents a class of similar activities, forming a compact representation of all the activities in the cluster. The activities represented by the final set of clusters are those that are modelled and recognized by the CASAS smart environment. It should be noted that currently, the number of clusters is provided to the clustering algorithm. However, alternative methods can be used to determine the number of clusters during runtime, by forming incremental clusters until no more change can be perceived.

RECOGNIZING ACTIVITIES

Once the activities are discovered, a model has to build for activity recognition. In our approach, Hidden Markov model is used to recognize activities from sensor data. Each model is trained to recognize the patterns that correspond to the cluster representatives found by ADM. A separate Markov model could be learned for each activity and the model that supports a new sequence of events would be selected as the activity label for the sequence.

For this task, hidden Markov model is used, which is a statistical model in which the underlying data are generated by a stochastic process that is not observable. HMMs perform well in the cases where temporal patterns need to be recognized. As with a Markov chain, the conditional probability distribution of any hidden state depends only on the value of a finite number of preceding hidden states.

An HMM model is specified, that using three probability distributions: the distribution over initial states, the state transition probability distribution and the observation distribution. The most likely sequence of hidden states are found that will be given to the observation in and by using the Viterbi algorithm [11].

One drawback of these HMMs sometimes it makes a very slow transition from one activity to another. To remedy this problem, an event-based sliding window is used and this limits the history of sensor events that the model remembers at any given time.

For activity recognition, a voting multi-HMM model is used as a boosting mechanism. Then multiple HMMs is constructed and recognize activities by combining their classifications using a voting mechanism. Specifically, the first HMM represents the first variation of all patterns (one hidden state per pattern), the second HMM represents the second variation of patterns, and so on.

The Viterbi algorithm is used for each HMM to identify the sequence of hidden states, one hidden state at a time, and then, using the described voting mechanism, then identify the most likely hidden state for the multi-HMM based on input from all individual HMMs. The multi-HMM is built automatically using the output of ADM's discovery and clustering algorithm.

PROPOSED WORK

Fuzzy Sate Q-Learning

To increase the prediction accuracy a Fuzzy-state Q-learning algorithm (FSQL) is proposed that is capable of learning a sequence of actions on the basis of the structure discovered by the process of DVSM. It is an extended version of the Modified Q-Learning Method with Fuzzy State Division, and aims to deal with the states under uncertain conditions, based on discovery of concurrent patterns from data. The procedure starts with some given fuzzy partitions. The number of fuzzy partitions decides the number of linguistic descriptions that are needed to reflect the model's complexity. So that the prediction accuracy is increased.

A. Fuzzyfication

First step in fuzzy logic is to convert the measured data into a set of fuzzy variables. It is done by giving value (these will be our variables) to each of a membership functions set. Membership functions take different shape. Different membership functions are used here is Triangular membership function, Trapezoidal Function, Gaussian membership function.

B. Fuzzy Rules and Inference System

The fuzzy inference system uses fuzzy equivalents of logical AND, OR and NOT operations to build up fuzzy logic rules. An inference engine operates on rules that are structured in an IF-THEN format. The IF part of the rule is called the antecedent, while the THEN part of the rule is called the consequent. Rules are constructed from linguistic variables. These variables take on the fuzzy values or fuzzy terms that are represented as words and modelled as fuzzy subsets of an appropriate domain.

C. Defuzzyfication

The last step of a fuzzy logic system consists in turning the fuzzy variables generated by the fuzzy logic rules into real values again which can then be used to perform some action. There are different defuzzification methods are Centroid Of Area (COA), Bisector Of Area (BOA), Mean Of Maximum (MOM), Smallest Of Maximum (SOM) and Largest Of Maximum (LOM).

The main advantages of using fuzzy logic system are the simplicity of the approach and the capacity of dealing with the complex data acquired from the different sensors. Fuzzy set theory offers a convenient way to do all possible combinations with these sensors. Fuzzy set theory is used in this system to monitor and to recognize the activities of people within the environment in order to timely provide support for safety, comfort, and convenience. Automatic health monitoring is predominantly composed of location and activity information. Abnormality also could be indicated by the lack of an activity or a abnormal activity detection which will cause or rise the home anxiety.

The first step for developing this approach is the fuzzification of system outputs and inputs obtained from each sensor and subsystem. These membership functions are ordered, firsly according to the area where they maybe occur and secondly according to the degree of similarity between them. The next step of our fuzzy logic approach is the fuzzy inference engine which is formulated by a set of fuzzy IF-THEN rules.

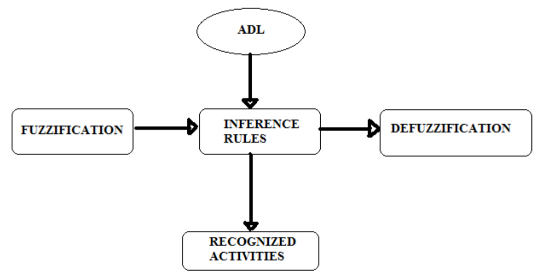


Figure.2 ADL Recognition Method

This second stage uses domain expert knowledge regarding activities to produce a confidence in the occurrence of an activity. Rules allow the recognition of common performances of an activity, as well as the ability to model special cases. A confidence factor is accorded to each rule and in order to aggregate these rules the Mamdani or Sugeno approaches are available under our fuzzy logic component. After rules aggregation the defuzzification is performed by the centroid of area for the ADL output. This framework also allows for rules to be added, deleted, or modified to fit each particular resident based on knowledge about their typical daily activities. This approach based on fuzzy logic provides robust and high accuracy recognition rate on the discovered data [12].

RESULTS

For discovering the activities in a smart environment, Discovering Frequent Discontinuous Activiti**es** method (3.1) finds the frequent patterns which are sequentially occurring in the smart environment and it eliminates the variant patterns.

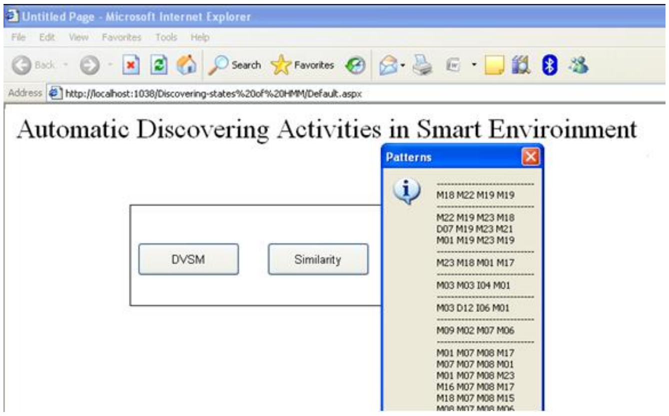


Figure.3 Intersted Pattern of activties

These interesting patterns are given to the Clustering Sequence into Groups of Activities method (3.2) which takes the Sensor ID and the Time as attributes for grouping the similar activities. Similarity is calculated based on mapping between the activities. These clustered activities are given to the Activity Recognition Method (3.3), it uses the HMM model which finds the probability between the activities by forming the states of activities.

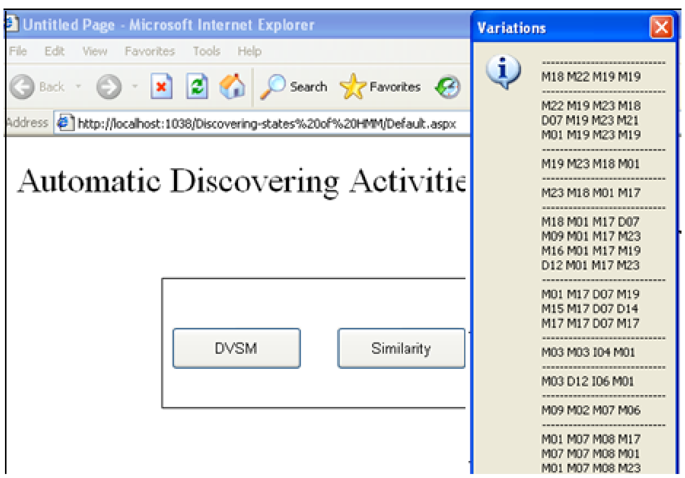
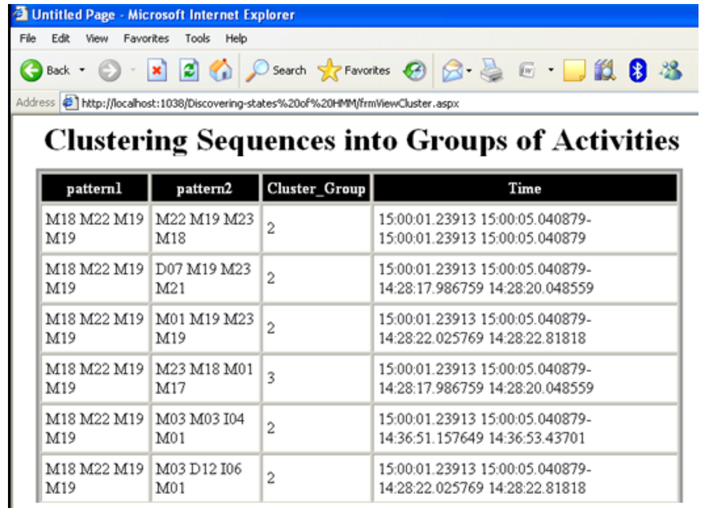
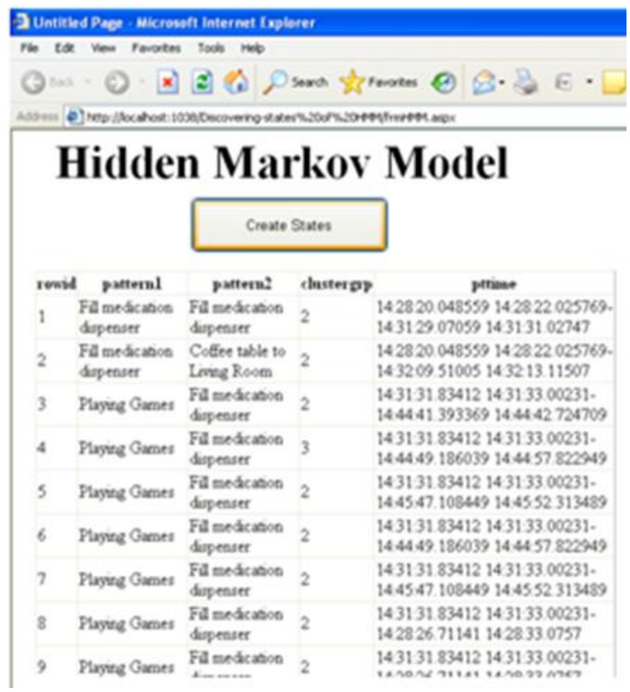


Figure.4 Pruned Patterns



pattern1	pattern2	Cluster_Group	Time
M18 M22 M19 M19	M22 M19 M23 M18	2	15:00:01.23913 15:00:05.040879-15:00:01.23913 15:00:05.040879
M18 M22 M19 M19	D07 M19 M23 M21	2	15:00:01.23913 15:00:05.040879-14:28:17.986759 14:28:20.048559
M18 M22 M19 M19	M01 M19 M23 M19	2	15:00:01.23913 15:00:05.040879-14:28:22.025769 14:28:22.81818
M18 M22 M19 M19	M23 M18 M01 M17	3	15:00:01.23913 15:00:05.040879-14:28:17.986759 14:28:20.048559
M18 M22 M19 M19	M03 M03 I04 M01	2	15:00:01.23913 15:00:05.040879-14:36:51.157649 14:36:53.43701
M18 M22 M19 M19	M03 D12 I06 M01	2	15:00:01.23913 15:00:05.040879-14:28:22.025769 14:28:22.81818

Figure.5 Clustered patterns



rowid	pattern1	pattern2	clustergp	pttime
1	Fill medication dispenser	Fill medication dispenser	2	14:28:20.048559 14:28:22.025769-14:31:29.07059 14:31:31.02747
2	Fill medication dispenser	Coffee table to Living Room	2	14:28:20.048559 14:28:22.025769-14:32:09.51005 14:32:13.11507
3	Playing Games	Fill medication dispenser	2	14:31:31.83412 14:31:33.00231-14:44:41.393369 14:44:42.724709
4	Playing Games	Fill medication dispenser	3	14:31:31.83412 14:31:33.00231-14:44:49.186039 14:44:57.822949
5	Playing Games	Fill medication dispenser	2	14:31:31.83412 14:31:33.00231-14:45:47.108449 14:45:52.313489
6	Playing Games	Fill medication dispenser	2	14:31:31.83412 14:31:33.00231-14:44:49.186039 14:44:57.822949
7	Playing Games	Fill medication dispenser	2	14:31:31.83412 14:31:33.00231-14:45:47.108449 14:45:52.313489
8	Playing Games	Fill medication dispenser	2	14:31:31.83412 14:31:33.00231-14:28:26.71141 14:28:33.0757
9	Playing Games	Fill medication	2	14:31:31.83412 14:31:33.00231-

Figure.6 States of Hidden Morkov Model

CONCLUSION

In this paper, ADL activities of smart home residents are tracked and recognized by using the Activity Discovery Method (ADM), which is used to discover the frequent patterns that naturally occur in the smart home. These frequent patterns are clustered for recognizing the activities. For recognizing the activities Multi HMM model is used, which is a supervised model and it has the difficulty while analyzing the complicated patterns. So, the Fuzzy State Q-Learning method will be used instead of HMM model to produce the effective result for recognizing the activities.

REFERENCE

[1] B.Reisberg, S. Finkel, J. Overall, N. Schmidt-Gollas, S.Kanowski, H. Lehfeld, F. Hulla, S.G. Sclan, H.-U. Wilms, "The Alzheimer's Disease Activities of Daily Living International Scale (ADL-IS)," Int'l Psychogeriatrics, vol. 13, no. 2pp. 163-181, 2001.

[2] V.G.Wadley, O. Okonkwo, M. Crowe, and L.A. Ross-Meadows, "Mild Cognitive Impairment and Everyday Function: Evidence of Reduced Speed in Performing Instrumental Activities of Daily Living," Am. J. Geriatric Psychiatry, vol. 16, no. 5, pp. 416-424, May2008.

[3] E.M.Tapia, S.S. Intille, and K. Larson, "Activity Recognition in the Home Using Simple and Ubiquitous Sensors," Pervasive Computing, vol. 3001, pp. 158-175, 2004.

[4] M.Philipose, K.P. Fishkin, M. Perkowitz, D.J. Patterson, D. Fox, H.Kautz, and D. Hahnel, "Inferring Activities from Interactions with Objects," IEEE Pervasive Computing, vol. 3, no. 4, pp. 50-57, Oct.-Dec. 2004.

[5] O.Brdiczka, P. Reignier, and J.L. Crowley, "Detecting Individual Activities from Video in a Smart Home," Proc. 11th Int'l Conf. Knowledge-Based and Intelligent Information and Eng. Systems (KES), pp. 363-370, 2007.

[6] T.van Kasteren and B. Krose, "Bayesian Activity Recognition in Residence for Elders," Proc. Third IET Int'l Conf. Intelligent Environments (IE '07), pp. 209-212, Sept. 2007.

[7] A.Ruotsalainen and T. Ala-Kleemola, "GAIS: A Method for Detecting Discontinuous Sequential Patterns from Imperfect Data," Proc. Int'l Conf. Data Mining, pp. 530-534, 2007.

[8] V.L.Levenshtein, "Binary Codes Capable of Correct Deletions, Insertions, and Reversals," Soviet Physics Doklady, vol. 10, no. 8, pp. 707-710, 1966.

[9] D.J.Cook and P. Rashidi, "The Resident in the Loop: Adapting the Smart Home to the User," IEEE Trans. Systems, Man, and Cybernetics, Part A: Systems and Humans, vol. 39, no. 5, pp. 949-959, Sept. 2009.

[10] J.A.Hartigan and M.A. Wong, "A K-Means Clustering Algorithm,"Applied Statistics, vol. 28, pp. 100-108, 1979.

[11] A.Viterbi, "Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm," IEEE Trans. Information Theory, vol. IT-13, no. 2, pp. 260-269, Apr. 1967.