Hierarchical Clustering Based Activity Tracking System in a Smart Environment

[1]Amitha. R, [2]K. Rajakumari
Dept. of IT, SNS College of Technology, Coimbatore, Tamilnadu, India

Abstract
The Healthcare industries have very great advantages of using a remote monitoring system. The machine learning and pervasive sensing technologies found in smart homes offer unprecedented opportunities for providing health monitoring and assistance to individuals experiencing difficulties living independently at home. In a smart home environment where the sensors are installed in important places. All the activity of the patients at home are monitored each and every second and that data is analyzed using the activity detection algorithm, which categorize the activities based on the set of records that are obtained from the sensor. When an abnormal activity is identified, then that information will be sent to the concerned care givers to get the immediate attention. Such cases like activities of the elderly people with Alzheimer's diseases are monitored. The system can deduce what a person is doing and act appropriately. The different set of activities that are categorized into Models. The Computer system constructs and maintains a model describing the environment. In normal activity detection algorithm, each model is defined based on the inputs that are passed on regular intervals. But in this activity tracking and mining approach based on Hierarchical Agglomerative Clustering algorithm (HAC), so that no input is needed for the model selection and the accuracy can be improved for each model.

Keywords
Activity Recognition, Data Mining, Sequence Mining, Clustering, Smart Homes

I. Introduction

A smart environment can be treated as an intelligent agent which perceives the state of the environment using sensors and acts upon the environment using device controllers. Researchers are recognizing that smart environments can assist with valuable functions such as remote health monitoring and invention. Smart environments have the potential to aid people with mental and physical limitations to provide resource conservation and to make our lives more comfortable and productive. Eating, dressing, cooking, drinking and taking medicine etc., are the individual's activity called Activities of Daily Living (ADLs). A variety of approaches have designed to model and recognize these activities. Individual's perform activities differently due to physical, mental, cultural & life style differences. There is a number of difficulties in this approach. First, it is due to the intersubject variability. Second, tracking only the preselected activities that ignores other important activities. Third, a significant amount of training data must be labeled and made available to the machine learning algorithm to track a predefined list of activities. In this paper introduced a supervised method of discovering and tracking activities in a smart environment that addresses the above issues. We also introduce clustering algorithm is proposed ,so that no input is needed to select the models and also create a naïve bayes classifier to represent the activities and their variations. The supervised nature of our model provides a more accurate approach for activity recognition than is offered by previous approaches, which take a unsupervised approach which is an automated approach produce some misidentification in classification because training data is not used for classification.Compared to traditional methods for activity recognition which solely utilize Naïve Bayes or other models for recognizing labeled activities, our approach first "classify" interesting patterns of activity, and then, recognizes these discovered activities to provide a more automated approach. We introduce a clustering method for discovering activity patterns, to group discovered patterns into activity definitions. Hierarchical agglomerative clustering algorithm is automatically recognize and detect the activities more accurately when they occur in the smart environment. In previous approach, they applied mining and clustering method together to discover the frequent patterns. The main limitation of this is user needs input for the model selection. For recognizing these activities boosted version of HMM model is applied. Here also one limitation called misidentification leads to less accurate model classification. These two limitations are overcome by our proposed approach in smart environment.

II. Realated Work

It is extremely difficult to find accurate and reliable detection of Activities of Daily Living (ADL). Researchers are now recognized the importance of applying health assistance and companies. There are different approaches for activity recognition. The approaches differ according to the type of sensor data that is used for classification, the model that is designed to learn activity definitions and the method that is used to annotate sample data.

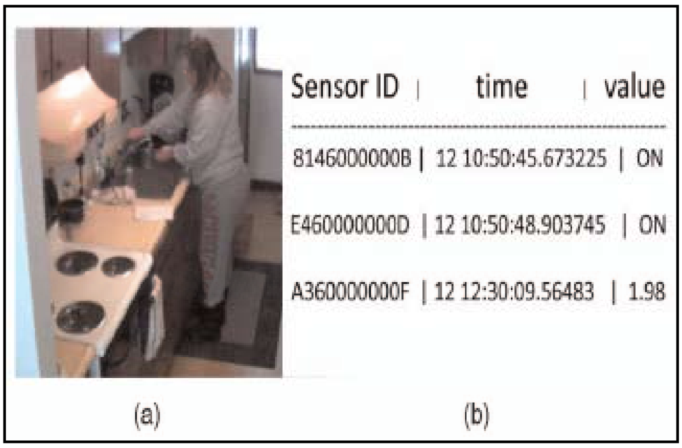

Fig. 1. (a) Resident Performing "Hand Washing" Activity. (b) This Activity Triggers Motion Sensor ON/OFF Events as Well as Water Flow Sensor Values

The sensor data is captured using a customized sensor network and each sensor reading is tagged with date, time of the event , and sensor value. For example figure 1 indicate different activities performed in a kitchen . After collecting the data from smart home, sensor data is annotated for activity recognition. A large number of sensor data events are generated, for that number of machine-learning models that have been used for activity recognition. In our approach, we apply a Naïve Bayes Classifier to classify the number of discovered patterns from sensor events. To annotate the sample data different methods are used. Here we take a supervised approach. Because, in earlier approach hand labeling is used that is very time consuming and also when activity monitoring is used for older adults with dementia, they cannot be expected to remember which activities they performed. So there may not be accurate information. In a supervised approach, discover activities that naturally occur frequently in an individual's home and then build models to recognize these activities. As a result no manual annotation of activity data.

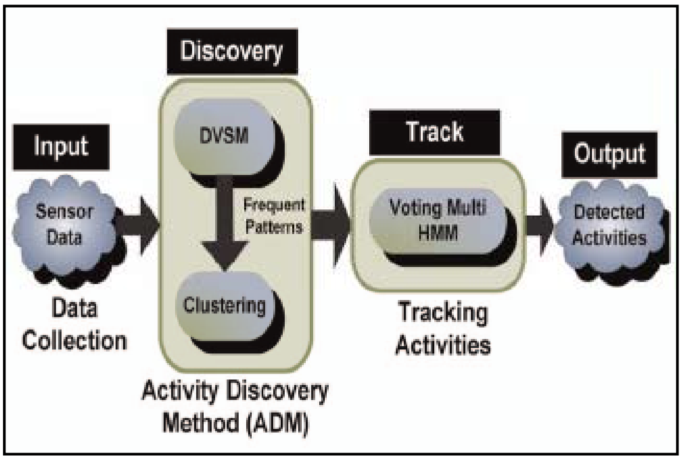

Fig. 2: Main Components of the Existing System for Discovering and Tracking Activities

In fig. 2, the main components of existing system is introduced. A new mining method called Discontinuous Varied Order Sequential Miner (DVSM) which is able to find frequent patterns that may be discontinuous and might have variability in ordering , so the issue is intrasubject variability and also employ activity clustering to group the patterns into activity definitions. Automatic determination of number of clusters present in the data set is a challenging problem. There are two aspects of clustering problem: finding the number of clusters and finding an appropriate partition of the data set. Here k-means algorithm is used for clustering. Next apply a HMM to represent and recognize those activities. As a result, it represents a fully automated approach to performing activity tracking to support smart environments but it produce less accurate results and also user needs to select the input for model classification. In our proposed approach we apply naïve bayes classifier for classification.Compared to HMM model , Naïve Bayes Classifier produce less error in model classification. Again it is combined with HAC algorithm ,so that it produced a more number of discovered patterns and models with more accuracy.

III. Classification of Activities

A. The Naive Bayes Classifier

The naive Bayes classifier is very popular in the data retrieval community, especially in textcategorization applications. A Naive Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem (from Bayesian statistics) with strong (naive) independence assumptions. A more descriptive term for the underlying probability model would be "independent feature model". In simple terms, a naive Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature. Depending on the precise nature of the probability model, naive Bayes classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for naive Bayes models uses the method of maximum likelihood; in other words, one can work with the naive Bayes model without believing in Bayesian probability or using any Bayesian methods. In spite of their naive design and apparently over-simplified assumptions, naive Bayes classifiers have worked quite well in many complex real-world situations. An advantage of the naive Bayes classifier is that it only requires a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification. Because independent variables are assumed, only the variances of the variables for each class need to be determined and not the entire covariance matrix. Fig. 3, represents our proposed Activity recognition system.

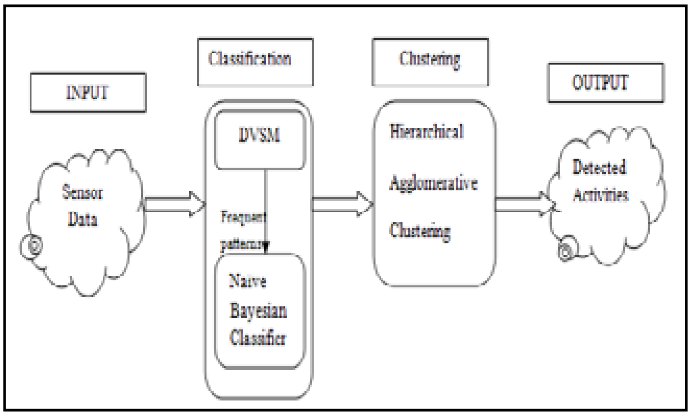

Fig. 3: The Block Diagram of the Activity Recognition System

IV. 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 represents the activities that we will model, recognize, and track. ADM uses hierarchical agglomerative clustering algorithm. Apply a hierarchical agglomerative clustering algorithm (HAC) to construct hierarchy. The HAC algorithm is a commonly employed classical hierarchal clustering algorithm. The result of HAC is a dendrogram representing the nested grouping of images.

The general HAC algorithm is as follows:

1. Put each image into a singleton cluster, compute a list of inter cluster distance for all singleton clusters, then sort the list in ascending order.
2. Find the pair of clusters with the most similar, merge them into one cluster and calculate the similarity between the new cluster and the remaining clusters.
3. While there is more than one cluster remaining, go to step 2, otherwise stop.

Based on the calculation of similarity between the non-singletons clusters a variety of hierarchical agglomerative techniques have been proposed. Single-link, complete-link and group average-link clustering are commonly used. In the single-link cluster the similarity between two clusters is the maximum similarity of all pairs of documents which are in different clusters. In the complete-link cluster, the similarity between two clusters is the minimum similarity of all pairs of documents which are in different clusters. In Group-Average-link clustering the similarity between two clusters is the mean similarity of all pairs of singletons which are in different cluster. The result of HAC algorithm requires no input for the model selection. So compared to other clustering algorithms, the accuracy of each model can be improved. The detected activities in form of images are again converted into video. So that each video is represented as different activity.

V. Result Analysis

Sensor data were collected continuously, resulting in some number of sensor events. We applied the activity discovery algorithms to these collected data. Sensor data is given as input, this data that consist of videos representing different activities in smart home. These videos are first converted into frames. Without user giving threshold value, The HAC algorithm automatically extracts the similar images into one cluster. In existing user need to specify the different number of activities to cluster, otherwise misidentification occur in building the model that represent the particular activity. So accuracy of the model reduced. After clustering the similar images together, bulid and extract the models representing similar activities .Each model is again converted into video that include the particular activity performed by an individual in smart home. HMM is easy to understand and erraritic in nature. Using Naïve Bayes classifier with HAC algorithm error rate is reduced. So that the overall accuracy can be improved by 98% to recognize the original activities.

Number of activities is detected from number of frames. Plot the graph between using HAC and naïve Bayesian Vs K-means and HMM. Number of discovered patterns will be high in Hierarchical clustering method and also using this method the accuracy of identification of each model can be improved. Fig. 4 shows the graph result of activity detection.

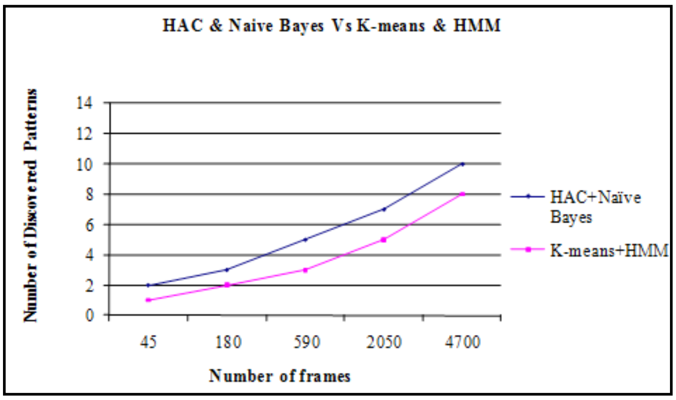

Fig. 4: Comparison Results of Graph

VI. Conclusions

In order to provide robust activity recognition and tracking capabilities for smart home residents, researchers need to consider techniques for identifying the activities to recognize and track. While most approaches target specific ADLs for tracking, this imposes a burden on annotators and residents and often introduces a source of error in the process. We introduce an alternative method for tracking activities in smart environments. In our approach, we employ our ADM algorithm to discover frequent activities that regularly and naturally occur in a resident's environment. Models are then learned to recognize these particular activities, and the resulting findings can be used to assess the functional well-being of smart environment residentsThis type of automated assessment also provides a mechanism for evaluating the effectiveness of alternative health. Interventions, and security alert systems.

The future work is to automatic classification of active learning methods can be extended to handle more sophisticated data types, such as graphs and sequences. One also can explore the effects of a noisy oracle and to propose methods for preventing the resulting accuracy degradation.

References

[1] Diane J. Cook, Lawrence B. Holder, Parisa Rashidi, Maureen Schmitter-Edgecombe,"Discovering Activities to Recognize and Track in a Smart Environment", IEEE Transactions On Knowledge And Data Engineering, Vol. 23, No. 4, April 2011.

[2] B. Reisberg, S. Finkel, J. Overall, N. Schmidt-Gollas, S. Kanowski, H. Lehfeld, F. Hulla, S.G. Sclan, H.-U. Wilms, K. Heininger, I. Hindmarch, M. Stemmler, L. Poon, A. Kluger, C. Cooler, M. Bergener, L. Hugonot-Diener, P.H. Robert, H. Erzigkeit,"The Alzheimer's Disease Activities of Daily Living International Scale (ADL-IS)", Int'l Psychogeriatrics, Vol. 13, No. 2, pp. 163-181, 2001.

[3] V. Rialle, C. Ollivet, C. Guigui, C. Herve,"What Do Family Caregivers of Alzheimer's Disease Patients Desire in Smart Home Technologies? Contrasted Results of a Wide Survey", Methods of Information in Medicine, Vol. 47, pp. 63-69, 2008.

[4] V.G. Wadley, O. Okonkwo, M. Crowe, 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, May 2008.

[5] S.-K. Noh, Y.-M. Kim, D. Kim, B.-N. Noh,"Network Anomaly Detection Based on Clustering of Sequence Patterns", Computational Science and Its Applications, pp. 349-358, Springer, May 2006.

[6] J. Yang, W. Wang,"Towards Automatic Clustering of Protein Sequences", Proc. IEEE CS Bioinformatics Conf., pp. 175-186, 2002.

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

[8] XindongWu, Vipin Kumar, J. Ross Quinlan, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda, Geoffrey J. McLachlan, Angus Ng, Bing Liu, Philip S. Yu, Zhi-Hua Zhou, Michael Steinbach, David J. Hand, Dan Steinberg,"Top 10 algorithms in data mining", Springer-Verlag London Limited 2007 (Kmeans)

[9] V.I. Levenshtein,"Binary Codes Capable of Correct Deletions, Insertions, and Reversals", Soviet Physics Doklady, Vol. 10, No. 8, pp. 707-710, 1966.

[10] K. Sequeira, M. Zaki,"Admit: Anomaly-Based Data Mining for Intrusions", Proc. Int'l Conf. Knowledge Discovery and Data Mining, pp. 386-395, 2002

[11] H. Saneifar, S. Bringay, A. Laurent, M. Teisseire,"S2mp: Similarity Measure for Sequential Patterns", Proc. Australasian Data Mining Conf. (AusDM), Vol. 87, pp. 95-104, 2008.

[12] Latifur Khan and Lei Wang,"Automatic Ontology Derivation Using Clustering for Image Classification1(HAC)".

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

[14] C. Dimitrakakis, S. Bengio,"Boosting HMMs with an Application to Speech Recognition", 2004

[15] Aaron S. Crandall, Diane J. Cook,"Coping with multiple residents in a smart environment".

[16] Pavan Turaga, Rama Chellappa, V. S. Subrahmanian, Octavian Udrea,"Machine Recognition of Human Activities: A Survey (HMM)".

[17] Abhinav Srivastava, Amlan Kundu, Shamik Sural, Arun K. Majumdar,"IEEE Credit Card Fraud Detection Using Hidden Markov Model(HMM)".

[18] C. Stauffer, W. Grimson,"Learning Patterns of Activity Using eal-Time Tracking", IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 22, No. 8, pp. 747-757, Aug. 2000.
[19] M. Philipose, K.P. Fishkin, M. Perkowitz, D.J. Patterson, D. Fox, H. Kautz, D. Hahnel,"Inferring Activities from Interactions with Objects", IEEE Pervasive Computing, Vol. 3, No. 4, pp. 50-57, Oct.- Dec. 2004.
[20] D. Cook, M. Schmitter-Edgecombe,"Assessing the Quality of Activities in a Smart Environment", Methods of Information in Medicine, Vol. 48, pp. 480-485, 2000.

AMITHA R received her B.E. degree in computer science and engineering from Dr. Mahalingam College of engineering, Pollachi, in 2008, the M.E. degree in computer and communication engineering, Coimbatore, Anna University, in 2012. Her research interests include Knowledge mining and engineering, cryptographic technique. At present, She is an assistant professor in SNS College of Technology, Coimbatore.