Tracking and Discovering Patient Regular Activities Using Clustering Technique

N.M. Nahas, D. Saravanan

Abstract—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. The actual data mining task is the automatic or semi-automatic analysis of large quantities of data to extract previously unknown interesting patterns such as groups of data records cluster analysis, unusual records anomaly detection, and dependencies association rule mining. In order to monitor the functional health of smart home residents, we need to design technologies that recognize and track activities that people normally perform as part of their daily routines, such as eating, dressing, cooking, drinking, and taking medicine. Although approaches do exist for recognizing activities, the approaches are applied to activities that have been preselected and for which labeled training data are available. In contrast, we introduce an automated approach to activity tracking that identifies frequent activities that naturally occur in an individual’s routine. With this capability, we can then track the occurrence of regular activities to monitor functional health and to detect changes in an individual’s patterns and lifestyle. In this paper, we describe our activity mining and tracking approach, and validate our algorithms on data collected in physical smart environments.

Index Terms—Data Mining, Machine Learning, Clustering, K-Means Cluster, Video Data.

1. INTRODUCTION

A convergence of technologies in machine learning and pervasive computing as well as the increased accessibility of robust sensors and actuators has caused interest in the development of smart environments to emerge. Furthermore, researchers are recognizing that 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 be able to complete Activities of Daily Living such as eating, dressing, cooking, drinking, and taking medicine[1][16]. Automating the recognition of activities is an important step toward monitoring the functional health of a smart home resident. When surveyed about assistive technologies, family caregivers of Alzheimer’s patients ranked activity identification and tracking at the top of their list of needs [2] [16]. In response to this recognized need, researchers have designed a variety of approaches to model and recognize activities. The generally accepted approach is to model and recognize those activities that are frequently used to measure the functional health of an individual [3] [16]. However, a number of difficulties arise with 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. This is certainly not always the case. For example, while individuals may regularly eat meals, they may go out to restaurants for the majority of their meals, which would make tracking this ADL challenging for a smart home. Even for an activity that is performed in the monitored environment, different individuals might perform it in vastly different ways, making the reliance on a list of predefined activities impractical due to the intersubject variability. In addition, the same individual might perform even the same activity in different ways, requiring methods that can also deal with intra subject variability. Second, tracking only preselected activities ignores the important insights that other activities can provide on the functional health of individuals [4]. For example, Hayes et al. found that variation in the overall activity level at home was correlated with mild cognitive impairment. This activity level was not restricted to predetermined activities but was related to the total activity level in the monitored environment. 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. Because individuals perform activities differently due to physical, mental, cultural, and lifestyle differences[5][16], sample data need to be collected and labeled for each individual before the learned model can be used reliably to track the individual’s activities and functional well-being. Unfortunately, collecting and labeling such sensor data collected in a smart environment is an extremely time-consuming task. [6] If the individual is asked to participate by keeping track of their own activities over a period of time, the process is additionally obtrusive, laborious, and prone to self-report error. In this paper, we introduce an unsupervised method of discovering activities in a smart environment that addresses the above issues. We implement our approach in the context of the CASAS Smart Home project by using sensor data that are collected in the CASAS smart apartment testbed. The unsupervised nature of our model provides a more automated approach for activity recognition than is offered by previous approaches, which take a supervised approach and annotate the available data for training [7].
Compared to traditional methods for activity recognition which solely utilize HMM or other models for recognizing labeled activities, our approach first “discovers” interesting patterns of activity, and then, recognizes these discovered activities to provide a more automated approach. We introduce a unique mining method for discovering activity patterns, along with a clustering step to group discovered patterns into activity definitions. For the recognition step, we create a boosted version of a hidden Markov model (HMM) to represent the activities and their variations, and to recognize those activities when they occur in the smart environment.

II. EXISTING SYSTEM

In general, activities are not always same. Means same activity taken place in different styles and methods (e.g. food taken in home and hotel is the wish of the own). So for recognition needs large amount of train data in then machine learning algorithm for each and every activity). Previous system monitors only preselected activities. The other activities done are not to be concerned. (i.e. if preselected activity is 5 (cook, eat, dress, wash, walk) then only those activities are recognized. The new activity not taken into consideration. Duplication can be occurred. When activities occur in at any interval. Predefined activities only defined, and it stored in to the data base. In general, activities are not always same, that is same activity takes place in different styles and methods. Each activity must be noted manually Manually noting down the activities is time consuming. If any new event occurs then it will be ignored. Previous system monitors only preselected activities. The other activities done are not to be concerned the new activity not taken into consideration.

III. PROPOSED SYSTEM

In this paper, the Use of unsupervised method of discovering or learning. Frequent sequential pattern mining algorithm can be used. Image processing Technique using chameleon algorithm. Duplication avoidance can be occurred in this system. And the advantages are Provides an automated approach for the activity recognition. Not like previous methods, proposed systems discover the interesting. Combination of the HMM label model, frequent mining method and clustering steps are going to recognize the activity. Discover activities that naturally occur frequently. Not like previous methods, like n manual annotation of activity data scripting of activities is required. Use of unsupervised methods of discovering or learning. The training sets are then compared with the live frames if it is similar then it will ignored. If any dissimilarities or new event occurs then it alerts the system. DVSM (Discontinuous varied order sequential miner) is used for duplication avoidance.[11]

IV. EXPERIMENTAL TECHNIQUE

Activity discovery method performs frequent sequence mining using DVSM to discover frequent patterns, and then, groups the similar discovered patterns into clusters. We use DVSM to find sequence patterns from discontinuous instances that might also exhibit varied-order events. As an example, DVSM can extract the pattern ha; bi from instances fb; x; c; ag; fa; b; qg, and fa; u; bg, despite the fact that the events are discontinuous and have varied orders. It should be noted that our algorithm is also able to find continuous patterns by considering them as patterns with no discontinuity. Our approach is different from frequent item set mining because we consider the order of items as they occur in the data.[14]. Unlike many other sequence mining algorithms, we report a general pattern that comprises all frequent variations of a single pattern that occur in the input data set D. For general pattern a, we denote the variation of the pattern as ai, and we call the variation that occurs most often among all variations of the prevalent variation, ap. We also refer to each single component of a pattern as an event (such as a in the pattern ha; bi). To find these discontinuous order-varying sequences from the input data D, DVSM first creates a reduced data set Dr containing the topmost frequent events. Next, DVSM slides a window of size 2 across Dr to find patterns of length 2.[15]. After this first iteration, the whole data set does not need to be scanned again. Instead, 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 (in the same iteration) to see if it is a variation of a previous pattern, or if it is a new pattern. To facilitate comparisons, we save general patterns along with their discovered variations in a hash table. To see if two patterns should be considered as variations of the same pattern, we use the Levenshtein (edit) distance to define a similarity measure simA; Bp between the two patterns. The edit distance is the number of edits (insertions, deletions, and substitutions) required to transform an event sequence A into another event sequence. CHAMELEON uses a graph partitioning algorithm to cluster the sparse graph data objects into a large number of relatively small sub clusters. It then uses an agglomerative hierarchical clustering algorithm to find the genuine clusters by repeatedly combining these clusters using the connectivity and closeness measures. CHAMELEON algorithm has been derived based on the observation of the weakness of two popular hierarchical clustering algorithms, CURE and ROCK. CURE and related schemes ignore information about the aggregate inter-connectivity of objects in two different clusters, they measure similarity between two clusters based on the similarity of the closest pair of the representative points belonging to different clusters. [12]. ROCK and related schemes ignore information about the closeness of two clusters while CHAMELEON uses k-nearest neighbor.
graph approach to represent its objects [13]. This graph captures the concept of neighborhood dynamically and results in more natural clusters. The neighborhood is defined narrowly in a dense region, whereas it is defined more widely in a sparse region.

V. EXPERIMENTAL SETUP

A. Creating sensor network
B. Data collection using discovery method
C. Activity recognition
D. Generation of new event.

A. Create sensor network

For detecting the activities we need sensors. The calculated number of sensors is fixed in the smart environment. The sensors are fixed in the smart environment that covers the whole activities in the smart environment. If suppose we consider the house is the smart environment then each and every portion of the house should covered by the sensors. The places like bedroom, bathroom, kitchen, storage and living room [Fig 1] [Fig 2]

B. Data collection using Discovery

The first step we must consider is how to identify the frequent and repeatable sequences of sensor events that comprise our smart environment’s notion of an activity. Once we identify the activity and associate specific occurrences of the activity, we can build a model to recognize the activity and begin to analyze the occurrences of the activity. By applying frequent sequential pattern mining techniques, we can identify contiguous, consistent sensor event sequences that might indicate an activity of interest. Many methods have been proposed for mining sequential data, including mining frequent sequences, mining frequent patterns using regular expressions, constraint-based mining, and frequent-periodic pattern mining. One limitation of these approaches is that they do not discover discontinuous patterns, which can appear in daily activity data due to the erratic nature of human activities [Fig 3] Another possible approach is to cluster the sensor events. Time series and sequence clustering algorithms have shown to be effective in constrained situations. Sequence mining algorithms have been successfully used in bioinformatics to discover related gene sequences. The limitation of clustering algorithms for our problem is that we do not want to cluster all of the data points, but only those that are part of an activity sequence which is likely to occur frequently and with some degree of regularity [Fig 6]. Build a model that will recognize future executions of the activity. Unique mining method is used for discovering the activity patterns Along with the mining method a clustering method also taken place for group the discovered patterns into activity definitions

C. Activity Recognition

Activity recognition in done by HMM model. This will allow the smart environment to track each activity and determine if an individual’s routine is being maintained.

The hidden Markov model is used to recognize activities from sensor data as they are being performed. Each model is trained to recognize the patterns that correspond to the cluster representatives found by ADM [Fig 4] Markov chains are used to recognize activities from sensor event traces that were segmented into non overlapping sequences [Fig 5]. A separate Markov model could be learned for each activity and the model that best supports a new sequence of events would be selected as the activity label for the sequence [Fig 7] [Fig 8]. For recognition hidden Markov model is used. (HMM) to represent the activities and their variations, and to recognize those activities

D. Generation of new Event

After build the trained event to check with existing trained activity model .If it’s not match then it will identified as new event otherwise it ignores as the trained event has not be done [Fig 5].

VI. EXPERIMENTAL RESULTS

A. Fig 1. Training Window

B. Fig 2. Training the video

C. Fig 3. Adding a Event
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 residents. While this is a useful advancement in the field of smart environment technologies for health monitoring and assessment, there is still additional research that can be pursued to enhance the algorithms. Currently, the user specifies a desired number of activities to cluster and model. In future work, we will investigate methods for automatically selecting this number based on the resident’s lifestyle. We will also investigate methods for seeding the clusters based on smart environment information and for incrementally modifying the patterns, clusters, and models as activities.
change over time. Ultimately, we want to use our algorithm design as a component of a complete system that performs functional assessment of adults in their everyday environments. This type of automated assessment also provides a mechanism for evaluating the effectiveness of alternative health.

VIII. FUTURE ENHANCEMENT

While this is a useful advancement in the field of smart environment technologies for health monitoring and assessment, there is still additional research that can be pursued to enhance the algorithms. Currently, the user specifies a desired number of activities to cluster and model. In future work, we will investigate methods for automatically selecting this number based on the resident’s lifestyle. We will also investigate methods for seeding the clusters based on smart environment information and for incrementally modifying the patterns, clusters, and models as activities change over time.

REFERENCES


AUTHOR’S PROFILE

N.M. Nihas Currently Doing his Master of Computer Application in Sathyabama University, Chennai, Tamil Nadu, India. His Area of Interest Data Mining, Software Engg, DBMS and Distributed Computing. He published Two National Conference and One International Conference in the area of Data Mining.