Human Activity Pattern Mining from Smart Home for Health Care

[1] P.Velvizhy, [2] M.Sairam, [3] B.Barath, , [4]A.Ajeeth Krishnaa
1,2,3,4 Anna University, Chennai, India

Abstract : We propose frequent mining and prediction model to measure and analyze energy usage changes sparked by occupants' behavior. The data from smart meters are recursively mined in the quantum/data slice of 24 hours, and the results are maintained across successive mining exercises. We also utilize the Bayesian network, a probabilistic graphical model, to predict the use of multiple appliances and household energy consumption. The proposed model is capable of short-term predictions ranging from next hour up to 24 hours and long-term prediction for days, weeks, months, or seasons. For the evaluation of the proposed mechanism, this method uses the UK Domestic Appliance Level Electricity dataset (UK-Dale) time series data of power consumption collected from 2012 to 2015 with time resolution of six seconds for five houses with 109 appliances from Southern England. It must be noted that in practice load disaggregation is carried by Non-Intrusive Appliance Load Monitoring (NALM) technique. NALM is a technique used to disaggregate a home's power usage into individual appliances and label them for further mining and analysis.

***IndexTerms* -Pattern mining, human activity recognition, healthcare, smart home.**

I. INTRODUCTION

Nowadays, there is an ever-increasing migration of people to urban areas. Health care service is one of the most challenging aspects that is greatly affected by the vast influx of people to city centers. Consequently, cities around the world are investing heavily in digital transformation in an effort to provide healthier ecosystems for people. In such a transformation, millions of homes are being equipped with smart devices (e.g., smart meters, sensors, and so on), which generate massive volumes of ne-grained and indexical data that can be analyzed to support smart city services. In this project, we propose a model that utilizes smart home big data as a means of learning and discovering human activity patterns for health care applications. We propose a model that utilizes smart home big data as a means of learning and discovering human activity patterns for health care applications. We propose the use of frequent pattern mining, cluster analysis, and prediction to measure and analyze energy usage changes sparked by occupants behavior.

II. RELATED WORK

Lately, there has been a growing interest in using smart home technologies for detecting human activity patterns for health monitoring applications. The main goal is to learn occupants behavioral characteristics as an approach to understand and predict their activities that could indicate health issues[3].

One approach uses Semi-Markov-Model (SMM) for data training and detecting individual habits and the other approach introduces impulse based method to detect Activity in Daily Living (ADL) which focuses on temporal analysis of activities that happen simultaneously[10]. Smart meters data are also used activity recognition using Nonintrusive Appliance Load Monitoring (NALM) and Dempster-Shafer(DS) theory of evidence. It collects pre-processed data from homes to determine the electrical appliance usage patterns and then employs machine learning-based algorithm to isolate the major activities inside the home. The issue is that the study has to perform two steps on the data to completely isolate the main activities[7][1].

Probabilistic detection uses everyday appliances usage from smart meter and smart plug data to trace regular activities and learn unique time segment groups of appliances energy consumption. The method employs hierarchical probabilistic model-based detection to infer about discovered anomalous behavior. This in tern can be used to understand the criticality of some abnormal behaviors for sustaining better healthcare[14][13][11].

The study aims to provide a portrait of activities of daily living for elderly patients independently living at home. The data is also used to mine important patterns of changes for short-term and long-term anomaly detection of urgent health conditions. The work uses Bayesian networks to predict occupant behavior from collected smart meters data. The study proposes behavior as a service based on a single appliance, but does not provide a model to be applied for real-world scenarios. It used time-series multi-label classier to forecast appliance usage based on decision tree correlations, however, the study takes only the last 24-hour window along with appliance sequential relationships[12][9][6][8].

This method suggests a clustering approach to identify the distribution of consumers temporal consumption patterns, however, the study does not consider appliance level usage details. This might not be applicable for human activity recognition since specic activities require individual and multiple appliance to appliance and time associations. The work considers the appliances ON and OFF status to detect usage pattern using hierarchical and c-means clustering. However, the study does not consider the duration of appliance usage or the expected variations in the sequence of appliance usage. The work proposes graphical model based algorithm to predict human behavior and appliance interdependency patterns and use it to predict multiple appliance usages using a Bayesian model. The above-discussed approach do not consider appliance level usage patterns, which is critical in determining human activity variations. Furthermore, our experiments are conducted using a much larger data set than existing studies although there are similarities in data analytics techniques between the proposed study and existing work[2][4][5].

III. PROPOSED MODEL

The block diagram for the entire system is shown in figure 1. The system represents the process which starts by cleaning and preparing the data and then applying frequent pattern mining for discovering appliance-to-appliance associations, i.e., determining which appliances are operating together. Then, it uses cluster analysis to determine appliance-to-time associations. With these two processes, the system is able to extract the pattern of appliance usage which is then used as input to the Bayesian network for short-term and long-term activities prediction. The output of the system is utilized by specific health care applications depending on the intended use. For example, a health care provider might only interested in knowing activities related to cognitive impairment where tracking the sequence of daily activities is crucial for reminding the patient when abnormal behavior is detected. Next subsection explain such processes and briefly outlines the theoretical background. The smart meter data in HDF file format are given as the input to the system. These data are converted into CSV format. The data in the file are then pre-processed and then pattern analysis(Basic pattern matching) in applied. The resulting data is then clustered by K medoid clustering using Partitioning Around Medoids(PAM). The clustered graphs are used to obtain association rules which is then used for Bayesian network prediction to obtain any anomalies. They are then intimated to the concerned.

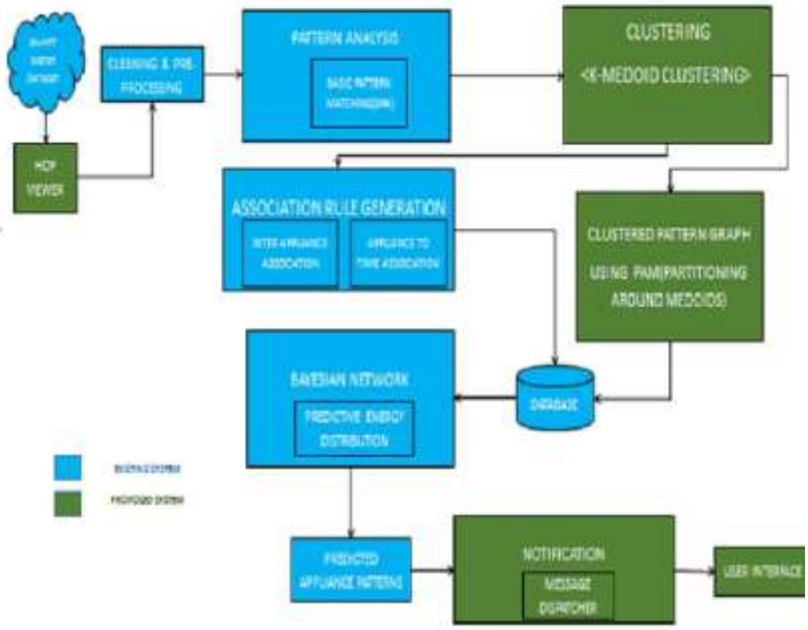

Figure.1 Block Diagram

A. CLEANING PRE-PROCESSING

We developed customized procedures to remove noises from the data and prepare it for mining. After cleaning and preparation, the data-set is reduced to 20 million as shown in figure 2. Smart meters time-series raw data, which is a high time-resolution data, is transformed into a 1-minute resolution load data; subsequently translated into a 30 minutes time-resolution source data, i.e. 242 = 48 readings per day per Appliances, while recording start time and end time for each active appliance users.

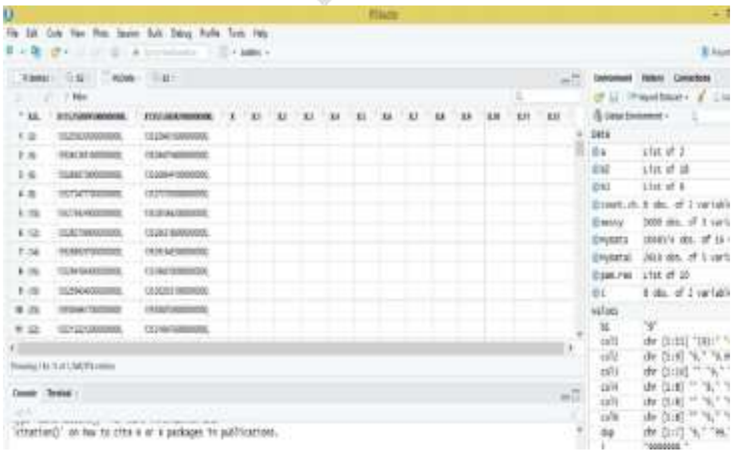

Figure.2 Representing Smart Meter Data in CSV format

B. PATTERN ANALYSIS(BASIC PATTERN MATCHING)

The aim is to discover human activity patterns from smart meters data. This method of pattern analysis as detects the patterns of these activities so that a health care application, that monitors sudden changes in patients behavior (e.g. patients with cognitive impairment), can send timely alert to health care providers. In pursuing such process, all appliances that are registered active during the 30-minute time interval are included into the source database for frequent pattern data mining.

The underlying concept of the model proposes pattern growth or FP-growth approach using depth-first divide-and-conquer technique. It exploits the benefits of pattern growth strategy and extends it to achieve incremental progressive mining of frequent patterns by mining in a quantum of 24 hours; i.e., frequent patterns are extracted from data comprising of appliance usage tuples for a 24 hour period, in a progressive manner.

C. CLUSTERING ANALYSIS(K-MEDOIDS CLUSTERING)

The underlying concept of the model proposes pattern growth or FP-growth approach is depth-first divide-and conquer technique. It exploits the benets of pattern growth strategy and extends it to achieve incremental progressive mining of frequent patterns by mining in a quantum of 24 hours; i.e., frequent patterns are extracted from data comprising of appliance usage tuples for a 24 hour period, in a progressive manner as shown in figure 3.

D. ASSOCIATION RULE GENERATION

Discovering appliance-to-time associations is vital to health applications that monitor patients activity patterns on a daily basis. In this section, a clustering analysis mechanism is used to discover appliance usage time. Appliance-to-time associations are underlying information in the smart meter time series data which include sufficiently close time-stamps, when relevant appliance has been recorded as active or operational. Using this data we can group a class or cluster of appliances that are in operation simultaneously or overlapping. The size of the cluster that describes such associations is deemed as the count of members in the cluster as well as its relative strength. Clustering analysis is the process of creating classes (unsupervised classification) or groups/segments (automatic segmentation) or partitions where members must possess similarity with one another, but should be dissimilar from the members of the other clusters. The distinct advantage of the clustering analysis is the non supervised nature of the process. We selected a 30 minute timespan/slice, for cluster segmentation, which will sufficiently capture the associations while minimizing the number of segments created; i.e., creating maximum 48 clusters for a day, whereas other clustering bases such as time-of-day, weekday, week and months have natural segmentation. For a dataset, DB, having n data points in Euclidean space. Partitional clustering distributes the data points from DB into k clusters, C1;C2; : : : ;Ck , having centroids c1; c2; : : : ; ck such that Ci DB, CiD ; and ci 6D cj for(1 i; j k).

An objective function based on Euclidean distance, distance(x; y) is used to measure the cohesion among data points, which rejects the quality of the cluster. This objective function is the sum of the squared k means algorithm seeks to minimize the SSE. And, make use of the silhouette score which is calculated based on the eluclidean distance to determine the optimal number of the clusters; i.e., k. In our model, incremental progressive clustering is obtained by consolidating existing and newly discovered clusters of each successive mining operation into the database. This incremental process is achieved by making sure all relevant cluster parameters such as centroid, SSE, Silhouette coefficient (width), data points and distance from the centroid are recorded in the database.

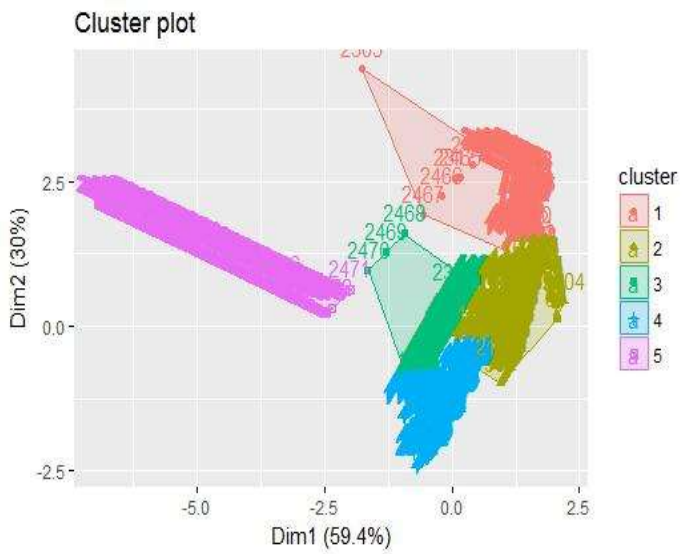

Figure 3. Result of Pattern Mining

E. BAYESIAN NETWORK PREDICTOR

In this section, we integrate the frequent patterns and appliance-to-time associations to learn about the use of multiple appliances and build the activity prediction model. The mechanism utilizes Bayesian network which is a directed acyclic graph, where nodes represent random variables and edges indicate probabilistic dependencies. One of the main features of a Bayesian network is that it includes the concept of causality. As mentioned above, our probabilistic prediction model is constructed based on integrating probabilities for appliances to-time associations in terms of hour of day (00:00 - 23:59), time of day (Morning, Afternoon, Evening, Night), weekday, week, month, season, and appliance-to-appliance associations.

The topology of the resulting Bayesian network has only one level of input evidence nodes, accompanied by respective unconditional probabilities, converging to one output node. The training data is derived from clustering and frequent patterns analysis where the probability of each appliance represents its operation during the specified period. This information is utilized by the Bayesian mechanism to determine and predict active appliances, operating simultaneously, using historical evidence from the cluster analysis (appliance time association) and frequent pattern mining (appliance-appliance association). Furthermore, appliance prediction results establish the foundation for human activity prediction from the next hour up-to 24 hours (short-term) and days, weeks, or months (long-term). We apply a Bayesian network for activity prediction based on individual and multiple appliance usage. This is significant for health applications that incorporate reminders for patients to perform certain activities based on historical data. For added accuracy of the system, the prediction model integrates probabilities of appliance-to-appliance and appliance-to-time associations, thus recognizing activities that occur in certain patterns more accurately

F. NOTIFICATION MESSENGER

This module broadcasts the newly found anomalies to other people concerned with the occupants.

IV. RESULT AND DISCUSSION

For the evaluation of the proposed model, we performed our experiments using the dataset UK-Dale along with the synthetic dataset to inspect intermediate and final results. The (UK-Dale) dataset includes time series data of power consumption collected between 2012 and 2015. The dataset contains time series data for five houses with a total of 109 appliances, having a time resolution of 6 seconds, from Southern England published by UK Energy Research Centre Energy Data Centre (UKERC-EDC). This dataset is one of the largest datasets having approximate half a billion records. Energy consumption measurement was conducted at appliance level using plug-in individual appliance monitors(IAMs).The underlying system for the proposed model is developed in R language.

The main objective of the experiments is to detect the appliance usage as an indication of human activity patterns and use the prediction model to forecast the short and long term activities inside the house. For a health care application, this means that our model can be used to feed mechanisms such as active monitoring, alert generation, health profiling etc.

Initially in the data set module, the data set is in HDF format(Hierarchical format). This data is then converted to ASCII and CSV format using Excel. The converted data is imported into R studio. The following algorithm analyses the frequent patterns in the dataset and produces a graphic representation. Thus the variation in the dataset based on the frequency of the patterns can be easily identified.

```
Pseudo code:
for all Frequent Pattern FP
do
        Search a frequent pattern FP in FPDB
        if Frequent Pattern found then
                Update frequent pattern in FP
        else
                Add a new Frequent Pattern to FP
        end if
end for
for all Frequent Patterns in Database FPDB
        increment Database Size by db24
end for
```

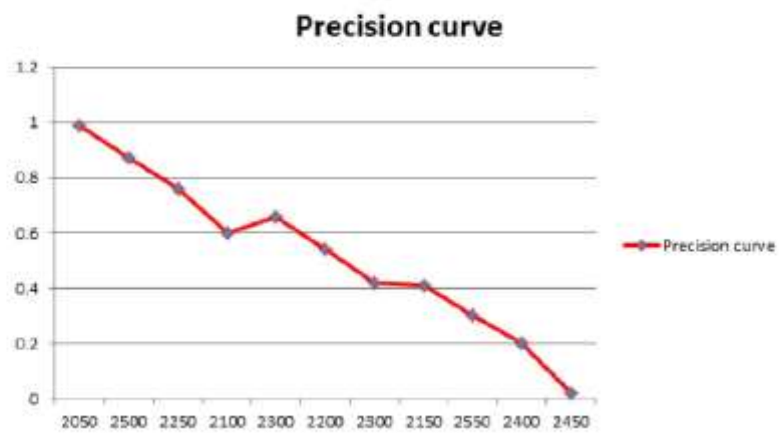

Figure 4. Precision Curve

The above graph plots the precision of variations in the meter readings over a 24 hour period.

The clustering module helps in identifying the variation in the graphic representation of the dataset based on frequent patterns. The variations are grouped into clusters based on the range of variation. The number of groupings can also be given based on the requirement.

Figure 5. Recall Curve

The above graph plots the recall of variations in the meter readings over a 24 hour period.

The prediction model utilizes appliance-to-appliance and appliance-time associations to predict multiple concurrent operating appliances. We can easily see the strong relationship between appliance usage inside the smart houses and human activity recognition. Learning the appliance-to appliance and appliance-to-time associations extracted from the frequent pattern mining and cluster analysis are key processes to track patients/people's routines and possibly provide them with health services when needed.

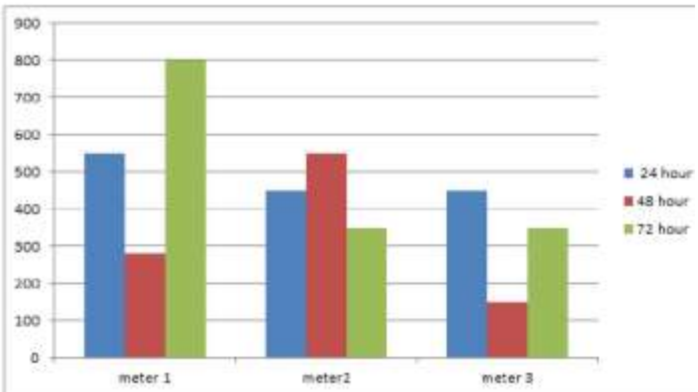

Figure 6. Accuracy Plot(actual data)

The above graph plots the accuracy of the variations in the actual meter readings over consecutive time periods.

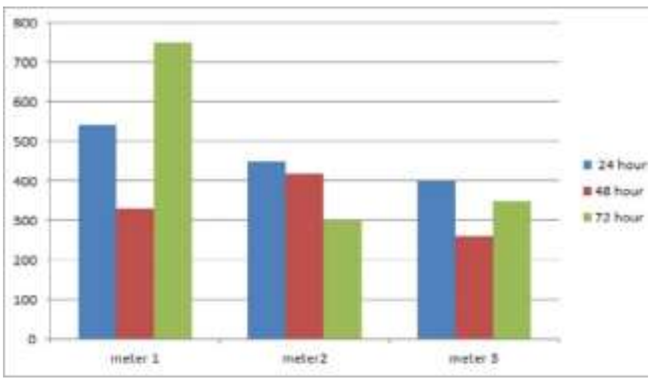

Figure 7. Accuracy Plot(after prediction)

The above graph plots the accuracy of the predicted variations in the meter readings over consecutive time periods.

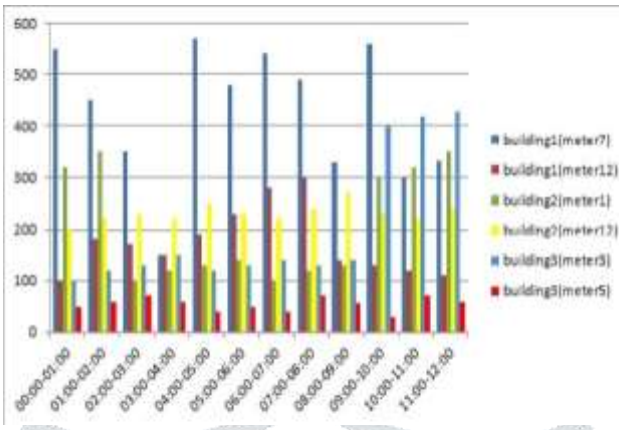

Figure 8. Comparison of appliances energy consumption among buildings

The above graph plots the variations of the meter readings compared to the other buildings over a 12 hour period.

V. CONCLUSION AND FUTURE WORK

In this project, we presented a model for recognizing human activities patterns from low resolution smart meters data. Occupants habits and behavior follow a pattern that could be used in health applications to track the well being of individuals living alone or those with self-limiting conditions. Most of these activities can be learned from appliance-to-appliance and appliance-to-time associations. We presented incremental frequent mining and prediction model based on Bayesian network. In our current work, through experiments, we found that 24-hour period was optimal for data mining, but we built the model to operate on any quantum of time. From the experiment results we have demonstrated the applicability of the proposed model to correctly detect multiple appliance usage and make short and long term prediction at high accuracy.

For future work, we are planning to renew the model and introduce distributed learning of big data mining from multiple houses in a near real-time manner. This will help health applications to promptly take actions such as sending alert to patients or care providers. Further more, we are planning to build a health ontology model to automatically map discovered appliances to potential activities. This means we can efficiently train the system and increase the accuracy of detecting human activities.

REFERENCES

[1] A. A. N. Shirehjini A. Yassine and S. Shirmohammadi, “Smart meters big data: Game theoretic model for fair data sharing in deregulated smart grids”, Elsevier Computers in Biology and Medicine, vol. 3, pp. 2743–2754, 2015.

[2] M.Mackay and P.Fergus C.Chalmers, W.Hurst, “Smart meter profiling for health applications”, Proceedings of International Joint Conference for Neural Networks, pp. 1–7, 2015.

[3] M.S.Hossain, “Cloud-supported cyber-physical localization framework for patients monitoring”, International Conference on Collaboration Technologies and Systems, pp. 118–127, 2016.

[4] M.S.Hossain, “A patients state recognition system for healthcare using speech and facial expression”, Journal for Medical Systems, vol. 40, pp. 272:1–272:8, 2016.

[5] M. S. Hossain and G. Muhammad, “Cloud-assisted industrial internet of things (iiot)enabled framework for health monitoring”, Computer Networks, pp. 192–202, 2016.

[6] J. Ploennigs J. Clement and K. Kabitzsch, “Detecting activities of daily living with smart meters”, Journal for Advanced Technology and Societal Change, pp. 143–160, 2014.

[7] L. Stankovic J. Liao and V. Stankovic, “Detecting household activity patterns from smart meter data”, Proceedings of the International Conference of Intelligent Environment, pp. 71–78, 2015.

[8] J.Kolte JC.Chelmis and V.K.Prasanna, “Big data analytics for demand response: Clustering over space and time”,Proceedings of the IEEE International Conference on Big Data, pp. 2223–2232, 2015.

[9] K.Jackand K.William, “The uk-dale dataset,domestic appliance-level electricity demand and whole-house demand from five uk homes”, Journal for Scientific Data, pp. 143–160, 2015.

[10] W.Abdul B.Song M.S.Hossain, G.Muhammad and B.B.Gupta, “Cloud-assisted secure video transmission and sharing framework forsmartcities”,Elsevier Computers in Biology and Medicine, pp. 430–437, 2017.

[11] S. V. B. Peddi A. Yassine P. Pouladzadeh, P. Kuhad and S. Shirmohammadi, “Mobile cloud based food calorie measurement”, Proceedings for the International Conference for Multimedia Expo Workshops, pp. 1–6, 2014.

[12] B. Y. Chiang S. Y. Peng Y. C. Chen, H. C. Hung and P. J. Chen, “Incrementally mining usage correlations among appliances in smart homes”, 2015.

[13] A. Yassine and S. Shirmohammadi, "A business privacy model for virtual communities", International Journal of Web based Communications, vol. 05, pp. 1–8, 2009.
[14] A. Yassine and S. Shirmohammadi, "Measuring users privacy payoff using intelligent agents", Proceedings of the IEEE International Conference on Computer Intelligence Measuring System, pp. 169–174, 2009.