Mining Human Activity Patterns from Smart Home for Healthcare Applications

G.Nivetha
UG Scholar, I.T
S.A.Engineering college.

J.Pooja
UG Scholar, I.T
S.A.Engineering college.

P.Senthil
Assistant professor, I.T
S.A.Engineering college.

Abstract- Big data that are collected from the smart devices have been used to retrieve the human activity patterns to improve the health status of the people, as there is a lot of financial investment in the digital transformation as an effort to provide the healthier ecosystem for people. In these transformation millions of smart devices are equipped around the home, which provides a massive amount of refined and sorted data that can be used to analyze the health patterns of the human. In this research, the work mainly focus on analyzing the big data extracted from human activities in the smart home for frequent pattern mining, cluster analysis, prediction to measure and analyze the energy consumption changes and patterns based on the appliance to appliance level and appliance to time association, which is completely related to human activities.

Keywords: *Smart Devices, Human Activity Patterns, Smart home, Digital Transformation, Cluster Analysis, Bayesian network.*

*****

1 INTRODUCTION

The need for the health care resources is being increased by the digital transformation. According to study, by the year 2050 digital transformation plays an important role. By advancement of these machines, a huge portion of data will be generated from smart devices [1]. For example, examining the changes of appliance usage can be used for indirect determination of the person's well being based on their historical data. Everyday activities reflects their regular habits, on observing their regular habits people's difficulties in taking care of themselves, like not washing his clothes, not using woven these help us to recognize any anomalous activities which might be an indication of ill health[4]. The interrelation between the appliance usage and routine activities is used by health care applications to detect potential health problems. This will automatically identifies the normal and abnormal behavior for independent living patients [7]. In a way, the large volume of data obtained from smart devices are analyzed to support health care services.

The model observes and analyzes the meter- readings of smart home equipment to identify the regular activities. Power consumption and the period of utilization are closely related to the resident's activities performed at household. For instance, if the "Oven" is ON, then it is most likely dealing with activity "Preparing Food". The time (e.g. morning or evening) of this operation may also indicate the type of the meal such as breakfast or dinner. Furthermore, people often perform more than one activity at the same time such as "Preparing their own Food","Watching programs over television", or "Listening to Music" which means multiple appliances are operated together at same time interval [4].

However, it is a challenging task since it is not so easy to detect usage dependencies among various appliances when their operation occurs at the same time [6]. Furthermore, tracing out an accurate human activity patterns and their prediction is greatly affected by the relationships between appliances usage and their time slots that have dynamic time intervals.

To handle the above-mentioned issue, the paper proposes frequent pattern mining and prediction model to measure and analyze energy usage changes observed in the household behavior. The data from smart meters are observed in the quantum/data portion of 24 hours, and the results are preserved across subsequent mining activities. Frequent pattern tree for mining the entire set for pattern recognition is observed based on the comparison between k-means clustering algorithm and DBSCAN clustering algorithm to identify the appliance-to- appliance cooperative from incremental mining of energy consumption data [12]. This is not only used to determine activity routines, but also, used for detecting sudden changes of human activities when utilized by health care application, that is required for attention of a health provider [13].

2. RELATED WORK

In this section, we review existing work in the literature, which employ smart homes data to analyze users' behavior. In paper [2] proposed a work on smart meters big data Game theoretic model for fair data sharing through a technique for sharing power consumption data in deregulated smart grids. This paper used the concept of differential privacy as an anonymity mean to minimize the leakage of information.

Similarly, the work in [3] adopted a dataset possessed from disparate homes. Smart homes contain very large number of meter readings by the users of the smart homes equipped by smart devices. The content of the meter readings varies from home to home based on the usage level of the equipment by the residents. This paper present an approach to assemble data from smart homes based on the usage of the appliances installed.A patient's state recognition system for healthcare using speech and facial expression is presented in [4]. It explains a model to address a overall framework on health care. It mainly deals with the concept of identifying a patient state for providing good recognition accuracy to provide low cost modeling. This paper mainly depends on two types of inputs considerably audio and video which are captured in a multi-sensory environment which showed an average detection efficiency over 98 percent.

3. METHODOLOGY

Block Diagram of proposed work

In this implementation, the household data is collected from the equipped smart devices fixed in smart home to measure the activity patterns of the living beings. After collecting the meter readings, the data is pumped for pattern mining using activity pattern algorithms.

A sequence of steps includes to carry out the work for detecting the human activity patterns. To implement the model, we collect the smart home data and apply pattern mining algorithms and clustering algorithms. The obtained result set is used to generate a trained Bayesian network, for classifying human activity patterns to predict abnormalities in the behavior of the smart home residents. First it starts by applying frequent pattern mining to discover appliance-to-appliance relations, that is for understanding which appliances are functioning together. Then, this model uses cluster analysis by comparing k means and DBscan clustering algorithms for deducing appliance-to-time associations. With the help of above two processes, the model is capable enough to deduce the pattern of appliance usage which is going to be used as input to the Bayesian network for activities prediction.

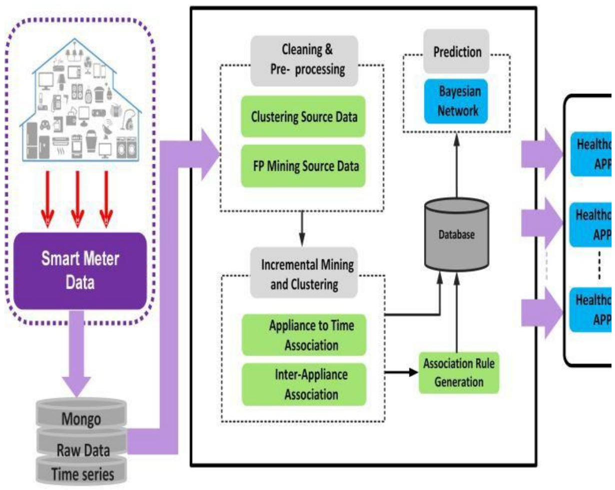

FIGURE 1. Model: Mining frequent patterns and activity predictions for health care applications in smart homes.

The output produced by the system is useful for specific health care applications depending on the anticipated use.

Steps of Proposed Work

- Smart homes are equipped with smart devices and smart meters.
- When the households use the smart devices, huge volumes of data are generated.
- Smart devices usage is like observing the routine activities of the households to recognize anomalous activities of the human.
- The huge volumes of data generated by the smart meter are collected.
- Then the data is classified based on the similarity of the data by using some algorithms.
- The data is clustered based on the given clustering algorithms.
- After the cluster analysis the source data is refined into supervised learning classification.
- Then the FP growth mining is applied to mine the source data.
- Then the obtained data is collected and set for the prediction to predict the human activity patterns based on the appliance usage.
- Bayesian network combine the frequent patterns and appliance-to-time associations to gain knowledge about the usage of various appliances at a time.
- Based on the Bayesian networks we build a activity prediction model which in terms helps us to predict the health conditions of the households.

Data Collection

The dataset used in this study is a collection of smart meters data from five houses in the United Kingdom (UK) in the initial stage of the cleaning process we remove noises from the data and prepared it for mining. We developed a artificial dataset for initial evaluation of the model, containing over 1.2million records in table I

Table I Ready-to-mine source data

Date	ST	ET	Active Appliances
2013-08-01	07:00	07:30	'2 3 4 12'
2013-08-01	07:30	08:00	'3 4 12'
2013-08-01	08:00	08:30	'2 4 12'
2013-08-01	08:30	09:00	'4 12'

ST = Start Time, ET = End Time, 2 = Laptop
3 = Monitor, 4 = Speakers, 12 = Washing Machine

Identifying Frequent Patterns in source data

As mentioned earlier, the aim is to discover human activity patterns from smart meters data. For example, activities such as "Watching TV, Cooking, Using Computer, Preparing Food and Cleaning Dishes or Clothes" are usually regular routines. Our aim is to detect the patterns of these activities so that a health care application, that monitors sudden changes in patient's behavior (e.g. patients with cognitive defects), can send timely alert periodically to health care providers. The energy detection of appliances (TV, Oven and Treadmill) is related to human activities such as leisure/relaxation time, food preparation, and exercising. Acquiring human activity patterns is not only observing the individual appliance operation, but also the appliance-to-appliance associations that is the patterns of activities that are combined such as washing clothes while exercising or watching TV. The underlying concept of the model is based on [10] which propose pattern growth or FP-growth approach and Apriori Algorithm for frequent pattern mining.

Cluster Analysis

Recognizing appliance-to-time associations has a major role in health applications that keeps a track on residents activity patterns all the time. In this section, to acknowledge about the appliance usage time clustering analysis technique is used. Appliance-to-time associations are underlying knowledge 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 defined 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.

K-Means Clustering Algorithm Partitioning the objects into commonly limited clusters (K) is done by it in such a manner that objects within each cluster persist as nearby as likely to each other.
Each cluster is considered by its Centroid i.e., its centre point. The partings used in clustering in most of the phases don’t really signify the spatial distances. In wide-ranging, the only resolution for this issue of conclusion global minimum is comprehensive choice of starting points. In a dataset, the appropriate number of clusters K and a set of k initial points, the K-Means clustering algorithm finds the predicted number of distinct clusters and their centroids. A centroid is the position where coordinates are acquired by means of computing each coordinates of the points, models assigned to the clusters.

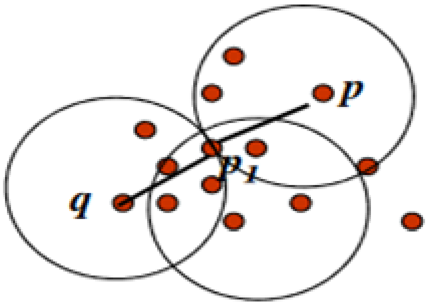

Figure 2 K-Means Clustering Points

DBscan Clustering Algorithm

The main impression of the DBSCAN algorithm is that, for each point of a cluster, the neighborhood of a given radius must contain a minimum number of points, that is, the density in the neighborhood must extinguish some predefined threshold. The Clustering procedure is based on the category of the points in the dataset as core points, border points and noise points and on the use of density link among points (directly density reachable,density-reachable, density-connected [Ester1996]) to form the clusters.

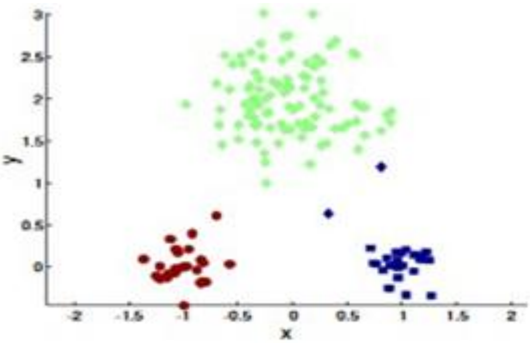

Figure 3 DBscan Clustering points

3.6. Bayesian Networks for Activity Prediction

In this section, we merge the frequent patterns and appliance-to-time associations to absorb about the use of multiple appliances and build the activity prediction model. The procedure utilizes Bayesian network which is a directed acyclic graph, where nodes illustrate random variables and edges indicate probabilistic dependencies. An example of Bayesian network, allege 6 random variables, is shown in Figure2. One of the main features of a Bayesian network is that it includes the concept of causality. For example, the link/arc between A to C in figure indicates that node A causes node C, which means that the directed graph in a Bayesian network is acyclic.

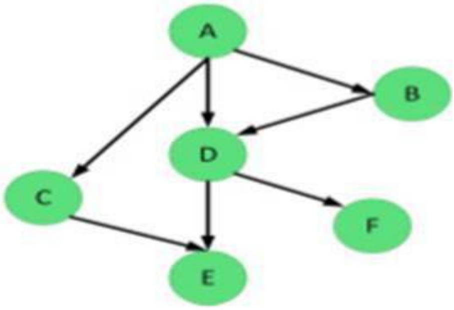

Figure 4 Bayesian Network

4. RESULTS AND ANALYSIS

In this section, K-Means clustering algorithm is compared with DBscan clustering. In addition, the usage of Bayesian networks for human activity pattern recognition also introduced.

Parameters	K-Means	DBscan
Approach	Partitional Based	Density based
Characterization	Centroid based	Dense Region based
Limitations	When clusters are of different Size, Densities, Non-globular shapes. When the data contains outliers.	Do not work efficiently when there are more number of clusters with different densities.
Advantage	More quicker compared to DBscan.	No cluster size is is demanded to form clusters.

Table II. Comparison between K-Means and DBscan

Smart data can be employed with either K- Means clustering algorithm or DBscan clustering algorithm. The selection of algorithm is mostly depending on dataset. If number of clusters is predefined and if the dataset is scalable K-Means can be applied. If no prior information about number of clusters then DBscan is feasible.

Here we used MOA tool to plot both K-Means and DBscan algorithms for a dataset. The results of both the clustering algorithms can be compared and based on our requirement we can use the dataset as a input for Bayesian networks.

Bayesian Networks for Activity Prediction

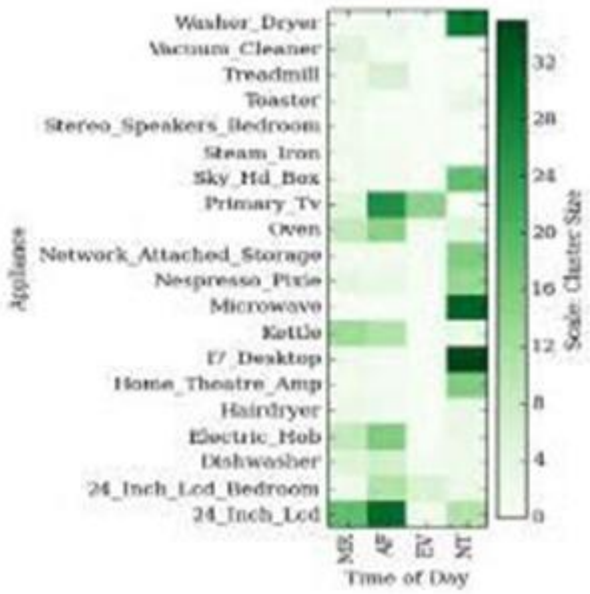

Figure 5 Augmental association of appliances- to-time of day

A Bayesian network is explaine by the probabilistic distribution presented in equation (1).

P(x1, x2, …,xn) = π p(xi|parents(xi))

Our probabilistic prediction model is raised based on integrating probabilities for appliances-to-time associations in terms of days, weeks, months, seasons and appliance-to-appliance level associations. The posterior probability for the contemplate work model is suggested in equation (2).

P(.) = p(Hour) x p (Time of day) x p (Week) x p (Week) x p (Month) x p(Season) (2)

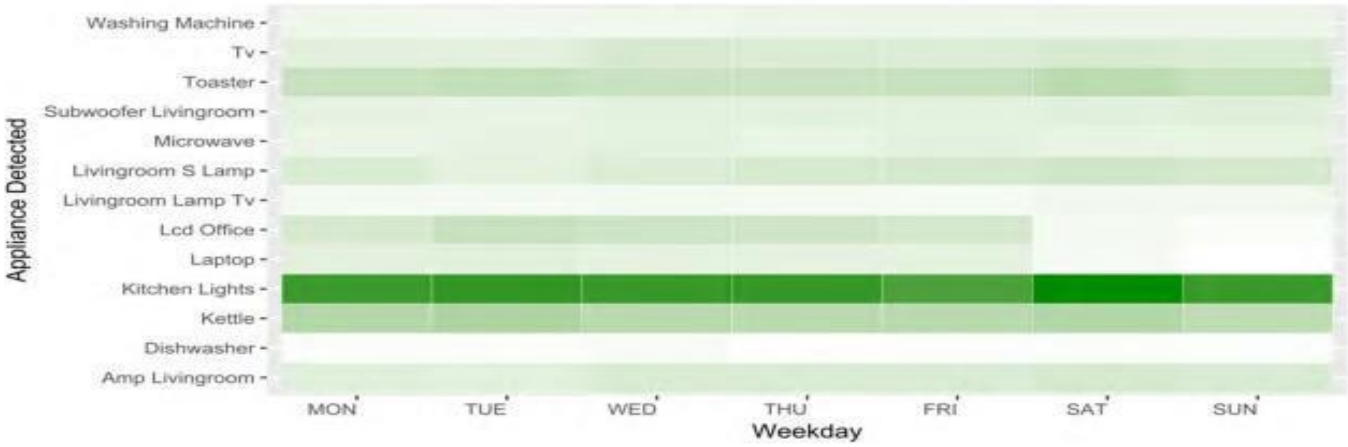

Figure 6 Appliance-Time association of the week

DISCUSSION

Most essential step in learning household activities is by mining associations of appliance usage. Figure 6 show the appliance-to-time association visible for time of the day and week

for respective house. We can clearly observe that between 2:30 and 5:00 Television, living area Lights are used simultaneously in the house with highest frequency during weekends.

Based on these truths we can observe the varying effect of days and weeks on the usage of appliances.

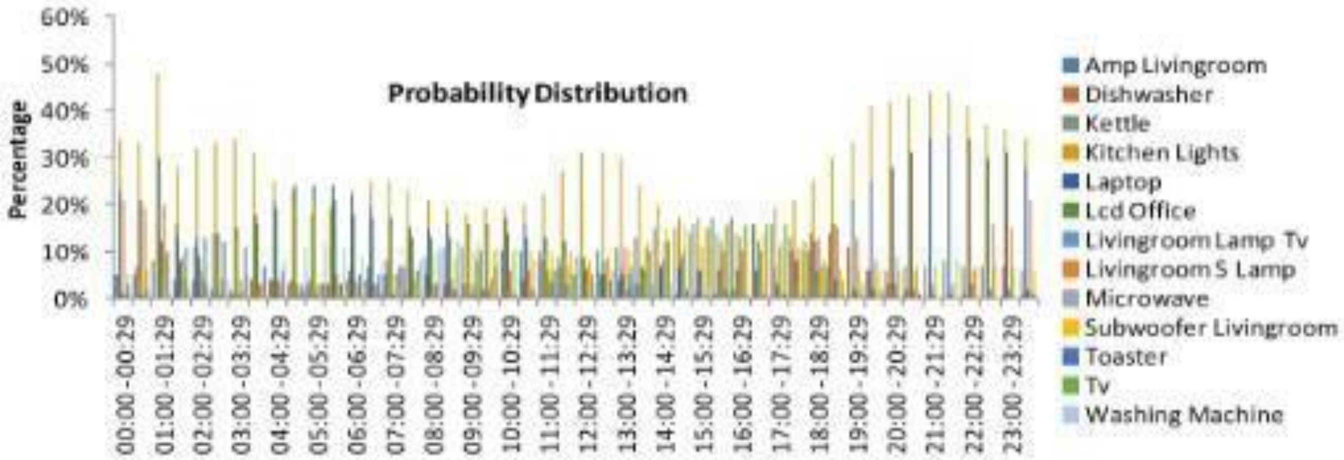

Figure 7 Probability Distribution of appliances for time of the day

Based on the probability distribution of appliances for time of the day in smart homes and probability distribution of appliances for time of week. Figure 7 and 8 show the probability distribution of appliance-to-time associations.

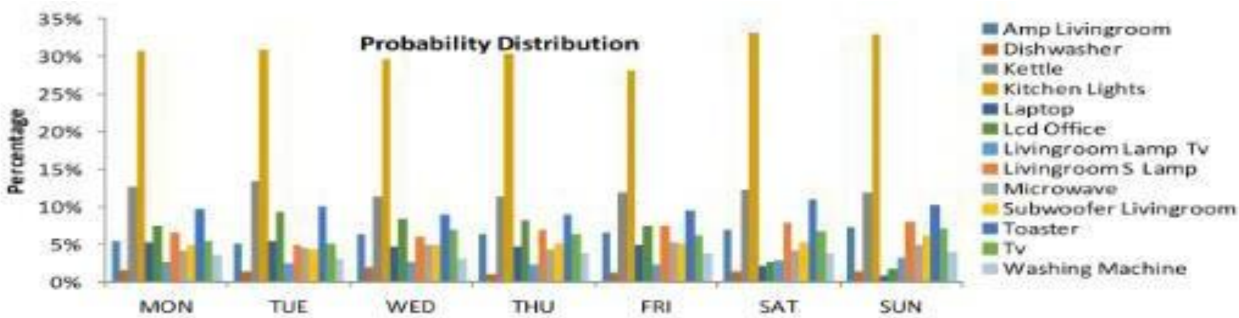

Figure 8 Probability Distribution of appliances for a week

The result shown in the figures represent the associations for 3 homes and it depends on processing 25% of the dataset. It is easily observed from appliance interrelation, behavior of inhabitant is observed, like to relax while preparing food.

5. CONCLUSION AND FUTURE WORK

In this paper, we presented a model for observing the human activity patterns from smart meter data and presented a model by comparing it with two most structured clustering techniques using the data. Most of the human activities can be observed from appliance-to-appliance and appliance-to-time associations. We presented a model by frequent pattern mining based on the clustering of the dataset and prediction model is presented based on the Bayesian network. In our current work, based on experiments, we found that a complete day period was optimal for data mining.

For future work, we are planning to reduce the model and introduce distributed learning of data mining from several houses in a instantaneous manner. This will help the households to be conscious and take actions by alerting the conditions of the households to care providers. Further, we can build a ontology model to automate the potential activities of the appliance which will increase the precision of perceiving human activities.

REFERENCES

[1] Abdulsalam Yassine, Shailendra Singh, and AtifAlamri," Mining Human Activity Patterns from smarthome big data for health care applications", IEEE Access, Vol. 5, 2017.

[2] A.A.N. Shirehjini, S. Shirmohammadi andYassine,"Smartmeters big data:Game theoretic model for fair data sharing in deregulated smart grids", IEEE Access,vol. 3, 2015.

[3] K. William and K. Jack, "The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from UK homes," Sci. Data, Sep. 2015.

[4] M. S. Hossain, "A patient's state recognition system for health care using speech and facial expression," J. Med. Syst., vol. 40, no. 12, Dec. 2016.

[5] M. UlAlam, N. Roy, M. Petruska, and A. Zemp, "Smart-energy group anomaly based behavioral abnormality detection," in Proc. IEEE Wireless Health (WH), Oct. 2016.

[6] K. Kabitzsch, J. Clement and J. Ploennigs "Detecting activities of daily living with smart maters", in Advance technology and Societal Change. Heidelberg, Germany: Springer, 2014.

[7] K. Basu, V. Debusschere, and S. Bacha, "Appliance usage prediction using a time series based classification approach," in Proc. IEEE 38th Annu. Conf. Ind. Electron. Soc. (IECON), May 2012.

[8] L. Hawarah, S. Ploix, N. Arghira, H. Joumaa, and K. Basu, "A prediction system for home appliance usage", Energy Buildings, vol.67, Sep. 2013.

[9] V.K. Prasanna, C. Chelmis, and J. Kolte, "Big data analytics for demand response: Clustering over space and time", in Proc. IEEE Int. Conf. Big Data (Big Data), Apr. 2015.

[10] T. Za_bkowski and K. Gajowniczek, "Data mining techniques for detecting House hold characteristics based on smart meter data," Energies, vol. 8, no. 7, 2015.

[11] Y. Yin, R. Rao, J. Pei, and J. Han, "Mining frequent patterns without Candidate generation: A frequent-pattern tree approach", Data Mining Knowl. Discovery, vol. 8, no. 1, 2004.

[12] J. Han, J. Pei, and M. Kamber, "Data mining: Concepts and techniques,"in Cluster Analysis: Basic Concepts and Methods, 3rd ed. San Mateo, CA, USA: Morgan Kaufmann, 2011.

[13] D. Heckerman, "Bayesian networks for data mining," Data Mining Knowl. Discovery, vol. 1, no. 1, 1997.