Online Health Monitoring Systemfor Patients in Smart Cities using IoT Applications

[1] B. Arunakumari, [2] K. Rohini
[1, 2] Assistant Professor
[1, 2] Department of CSE
[1, 2] AITS-Rajampet, A.P, India
[1]*bayanaaruna14@gmail.com*, [2]*rohini.tdpr@gmail.com*

***Abstract*—**In current situations people's migration from the villages to the cities became common. Healthcare services are known well by the people in corporate world to maintain good health habits who are feeling sick. At present the enhancements in technological aspects proceed with the IoT in smart homes, for instance automation of the electronic gadgets by smart meters or sensors. Even though there is having many enhancements in those sectors, limited research is going on towards healthcare. The sensors or smart devices will capture the daily activities of a patient who lives in smart home for their health monitoring. We propose the software system that can capture the patient's activities and transfers this data to a doctor by taking the state of various appliances of smart home into consideration. The essential frequent patterns and the association rules are retrieved from sensor information and this will be used for forecasting the patient's behavior. Our proposed works are; initially the daily activities are gathered by the sensor at a regular interval with a particular threshold. The association rules are extracted by using frequent pattern mining from such relational data. K-means clustering is applied for that resultant data. Eventually we used a Bayesian network model which is used to predict a patient's behavior and also to manage the health habits remotely by the supervisor.

Keywords – *IoT, Big data, Patient Health Management, classification.*

I. INTRODUCTION

The everyday interaction between the ubiquitous sensors and gadgets to connect the things physically and logically via virtual network is known as Internet of Things (IoT). To make the new procedure "anytime, anywhere any service for anybody". The IoT includes with the various techniques. There are having some probabilities produced by the IoT which made this possible to afford several related applications. From thosesmart homes are highly improved with the motto of users comfort and providing assurance for their security and with least operational costs as its criteria. As smart home is an automated environment it can do monitor, identify and record the daily activities with the utilization of various sensors and the communication technologies. The user's everyday activities develop patterns that perform a crucial role in a smart home environment [1, 2].

The identification of user's everyday activities monitored remotely. It has applications in different fields like health care and daily care [3]. Therefore the research upon identifying home activity is attaining high interest, particularly on the basis of recent trends to transfer healthcare from hospitals to the patient's home and allow them to lead independent life. Anyhow some possible applications go far beyond healthcare. For instance finding daily activities will helps in giving support to the home automation and saves energy in the smart homes. Recognition of home activity relies on inference making through data fusion from various sensors and uncertainty relevance because of stochastic nature of user behavior and poor sensing equipment. Some of the instances of finding activities in smart homes work with D-S theory. Lee et al. proved that D-S theory of evidence will provide a way to incorporate and reduce the impact of uncertainty efficiently. The D-S theory of evidence along with the lattice structure is used to find basic activities such as toilet flushing, for assistive living purpose. In this paper we consider many important routine activities are detected only with the use of energy monitors as smart meter information. By 2030 every home in India will be equipped with a smart meter, and with the same trends in the rest of developed nations. Since each home equipped with the sensors for finding activities so there is no need to have extra hardware. The remaining paper is framed as follows. Section 2 provides brief description on related and literature study of our work. Section 3 depicts our proposed work. Section 4 describes the results and discussion. Conclusions are drawn in Section 5.

II. RELATED STUDY

Many researchers proposed IoT based smart home environment to improve the security and user's comfort with the least operational costs. It is clear that the usage of sensors in a smart home is indispensable to keep track of

user activities. These user activities are monitored and the frequent patterns are created on the basis of user's location and their corresponding operations upon the appliances.Additionally other researchers are used to reduce the risks regarding activity detection with some varied approaches in many real world activities.

In [5], A EM-algorithm helps to group the similar objects. It is easy and faster in terms of performance but the number of features and objects varies its efficiency. In [7] the hierarchical clustering paradigm is utilized in distributed environment computes its performance and preciseness with the application of valid measures like entropy, time and coefficient of time. It is not required to mention number of clusters formerly and implementation is easy. In spite of the hierarchical algorithm provides clusters with poor quality and takes more time to execute when the large data set is given. A SOM algorithm [8] yields the more accuracy while classifying the objects into related clusters. Besides it gives better results over k-means and EM-clustering algorithm while random datasets are utilized. As 'k' (number of clusters) increases automatically the performance of it decreases otherwise by using large dataset, this paradigm gives poor results.

In 2015 C. Chelmis, J. Kolte and V. K Prasannaintroduced their research related work related to big data analytics for demand response: Clustering more time and space. Our paper depicts the usage of several data representations upon electricity consumption. It is also explained the use of behavioral patterns identified at (i) different times in a day, (ii) days-of-the-week or (iii) per annum for a clientele and similar ones are mined in clientele's with different features exact clustering of time-series data.

In 2015 K. Gajowniczek and T. Zabkowski proposed their work related to data mining approachesto identify the household characteristicsbased on smart meter information. Here the goal is to regulate the usage patterns of household's structure;hence by putting the view of utilization of smart meter systems it givesextra knowledge on its usage. Many of unsupervised machine learning methods is processed during the observation of the pattern's usage bymany households. This work produces the related outputs to the smart meter systems that contribute to highly developed energy consciousness; it is suitable to forecast the exact utilization and gives input to the demand response systems in households and suggests the users on energy saving sequentially.

2.1.2. Clustering

Clustering is a quite common theoretical task. Here anybody is able to find the group of finite clusters to depict the data.RuiVelosoa[9] implemented vector quantization for clustering purpose to forecast the readmissions in intensive medicine. This procedure used the paradigms such as k-means and k-mediods and so on. What are all the data sets utilized for this study are attained from the patient's diagnosis process and as well as the laboratory reports. By using Davies-Bouldin Index the evaluation of every algorithm is done. The algorithms such as k-means, x-means and k-mediods produce better,fair and poor results respectively. Depending on these results the researchers choose the good result that characterizes the various patients with the greater probability of a readmission.In this paper we concentrated only on the method of vector quantization.

2.1.3. Classification

Classification is nothing but a procedure for finding a predictive learning function which performs the classification of a data item into a predefined class that are not same. The following are the subsections that cover the classification related work.

2.1.4. Decision Tree

Most of theliterature explored the models of the decision tree for analyzing the clinical data.The researchers Sharma & Om [10], Wang et al. [11] and Zolbanin et al. [12] utilized the decision tree procedure for their research work purpose. For forecasting it is essential to examine data and make a tree and the rules of it are used. All those works noticed above uses a decision tree for the dataset to increase the performance in terms of accuracy. For this research work a balanced dataset is used.

2.1.5. Bayesian Classifier

The Bayesian classifiers are well known for its computational efficiency. Those are able to handle the skipped data efficiently. With this benefit the researchers can get better accuracy by the created models.Since the models are implemented by the use of Bayesian classifiers and also proved that model is appropriate because the average rate of the model tends to increase the accuracy in forecasting. This strategyis better only if dataset has the skipped data. The brief description about Bayesian Network is as follows. It is a directed acyclic graph, in which the nodes denotes the interest variables and the edges represents the association among those variables [Pearl 1988].

III. PROPOSED WORK

The flow of our proposed work is shown in Figure 1. It starts by the data pre-processing such as data cleaning and data preparation and then we have applied the FP mining to identify the association among different appliances. With this we can know which appliances are working together. Here we used cluster analysis to find the association between the appliance and time. After these steps now the system can get the patterns of appliance and this can be given as an input to a Bayesian Network for short term and long term prediction. The output of framework is used in healthcare applications relying on the intended use. We depicted the used theory concepts of approaches in brief in the next section.

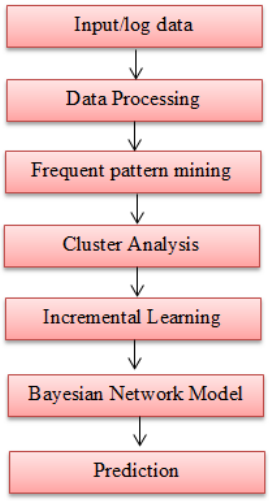


Figure 1: Flow of proposed work

3.1 Data Preprocessing

Sometimes the data pre-processing is neglected but in the case of data mining process it is not.If our dataset consists of irrelevant and a noisy data, it is hard to do knowledge discovery process over the data training. The quality of input data influences final outputs. For getting exact results eventually, the data is to be preprocessed efficiently and we may utilize data cleaning, integration, transformation and data reduction to make this process easy. Figure 2 shows the stages. For our work we collected the attribute values, and so there will be no null values and we chose the stipulated time and that yieldsexact results because of consistency. We are collecting the patient's data and daily activities by including various appliances in a smart home. For the appliances we took the binary values. For instance take 'Television' and for this the value '1' denotes its 'ON' state and '0' denotes its 'OFF' state.

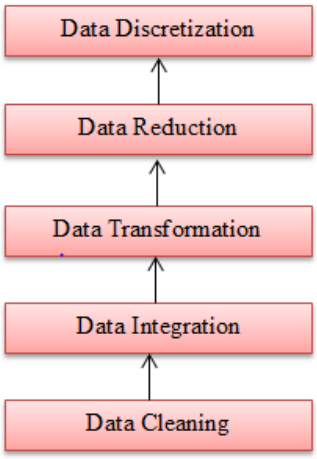
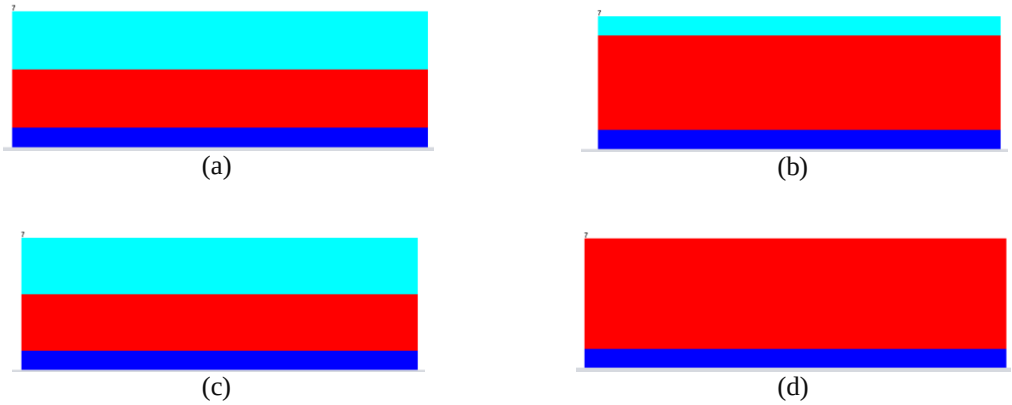


Figure 2: Data Pre-processing stages

3.2 Mining Frequent Patterns from Log Data

Our work primary goal is to identify the patterns on human activities basis from the smart meter data. There exists several activities consider that 'Watching Television', 'Using Laptop', 'Using Micro oven' etc.We must detect the frequent patterns from existing activities, which are essential in the healthcare applications that monitor the upcoming changes in the behavior of the patient.The doctor is able to see log data and suggest the measures for the patient's on the basis of formed association rules. Most essential if a patient is taking rest for a longer period, a doctor will suggest the patient on association rules basis and is very important for the heart patients.We can set different time interval 30-60 minutes; the patient can upload the smart meter data and the frequent patters are formed accordingly.The numbers 1,2 and 3 denotes the various daily activities. We have chosen this concept from [24], [25] that use the FP-growth approach by using divide and conquer method.Let $\{I_1, I_2, \text{and } I_3 \ldots I_n\}$ are the item sets that are having 'n' number of items. I_k denotes the 'k^{th}' item in the item set I_n.



(a) (b) (c) (d)

Figure 3: Result of Numeric to Binary Filter for our data set. (Only four attributes are shown in (a), (b), (c) and (d)).

3.3 Cluster Analysis

It is crucial for healthcare applications to mention the relationships between time and the appliances. It becomes simple while monitoring the patient's everyday activities on a timely basis. We used the cluster analysis to determine the utilization of an appliance on a timely basis. Suppose if a patient is watching a television in a specific period of time on whole day (00:00-23:59) and this activity is monitored by time of a day in the three slots (i.e. morning, afternoon and evening). The association between an appliance and the time on the basis of collected data through smart meter. For our work instead using smart meter we used sensor. The data that is recorded will be stored in the database and also we can combine them as a set of products so that a patient is using it in a specific time. Whenever we upload the recent activity, the incremental clustering is performed and later the formation of clusters is done accordingly and the below Figure 6 represents the corresponding clusters. Finally we integrate the recognized patterns and the association between the time and appliances and are provided to the machine learning model called as a Bayesian Network model for activities prediction. Low is '0', Over is '1'

Attribute	Full Data (7.0)	0 (6.0)	1 (1.0)
DATE	5/20/2018	5/20/2018	5/20/2018
SHIFT	Evening	Evening	Morning
NO.OFMINUTES	60	60	60
WASHING MACHINE	0.1429	0	1
BOWL WASHER	0.1429	0	1
THREAD MILL	0.2857	0.1667	1
YOGA MAT	0.1429	0	1
MICROWAVE OVEN	0.2857	0.3333	0
NESPRESSO	0.1429	0.1667	0
TV	0.1429	0	1
LAPTOP	0.5714	0.5	1
WORK	0	0	1
FOOD	0	0	0
WORKOUT	0	0	1
REST	0	0	1

Figure 4: Summary of Cluster analysis

As depicted in earlier section BN is the graph that is having nodes and the edges. The nodes represent the random variables whereas the edges represent the probabilistic dependencies. We have chosen this model on its key feature "causality". The sample Bayesian network model comprising '6' nodes is shown in below figure 5.

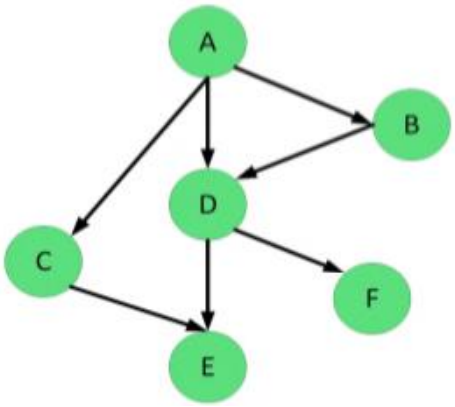


Figure 5: Bayesian network with '6' nodes.

IV. RESULTS & DISCUSSION

This section covers our experimental study and the results of our proposed work. Our software system is capable of storing huge amount of data base and can also be applied big data frame work for getting better performance and computations. The patients are able to register through the network and the details are stored in cloud or any other information repository. A list of doctors of different categories is maintained in our database, for example to give precautions or remedy for various diseases. Once the doctor accepted the request from the patient a connection is established between the patient and doctor and is able to exchange the messages in online. Along with this, the doctor is now able to monitor the daily activities done by the patient in the smart home through our software system. For example, the doctor can send a warning message or a general message to the patient to give precautions to the patient once he observed the activity by the patient, this is done by the sensor device over IoT. But in our work, the patient can upload his daily activities from his login/family login. This information can now transferred to the doctor and patients are able to get the suggestions from the doctor in online. We used our algorithm for extracting the frequent patterns/ associations from the captured data from the patient at the doctor side. As in our work, we used to upload a data file or binary information by the patient to give input for our algorithm, but we can getter better user satisfaction once the data is collected automatically be the smart meter or sensor data using IoT. The input for the algorithm is shown in Figure 6, the details uploaded by four patients along with their log information. Clustered data is shown in Figure 7. The cluster results may vary once, the data has been uploaded by the patient, since the activities of the patient of the other time slot may be different with the current clustering result. The result is shown in Figure 9. We also shown the list of work out, work, rest and the food consumed by the patient in Figure 10.

WM	BW	TM	YM	MO	NES	TV	LT
0	0	0	0	1	0	0	1
1	1	1	1	0	0	1	1
0	0	0	0	0	0	0	1
0	0	0	0	0	1	0	1
0	0	0	0	1	0	0	1
0	0	0	0	0	0	0	1
0	0	1	0	0	0	0	0

Figure 6: Activities of the patient with unique appliances

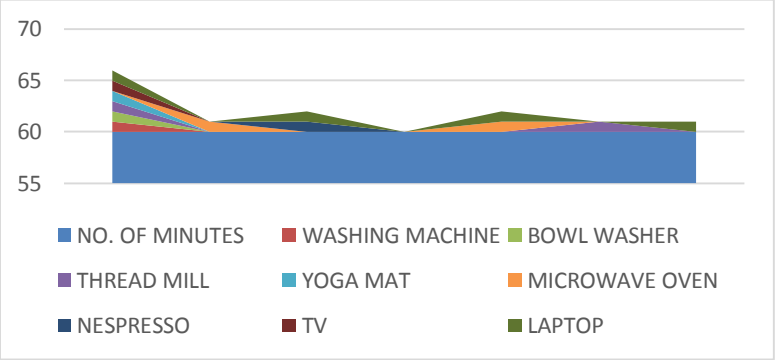


Figure 7: Cluster representation of the appliances

WORK	FOOD	WORKOUT	REST
OVER	LOW	OVER	OVER
LOW	AVERAGE	LOW	LOW
LOW	AVERAGE	LOW	AVERAGE
LOW	LOW	LOW	LOW
LOW	AVERAGE	LOW	AVERAGE
LOW	LOW	AVERAGE	LOW
LOW	LOW	LOW	AVERAGE

Figure 8: Result of workload by the patient

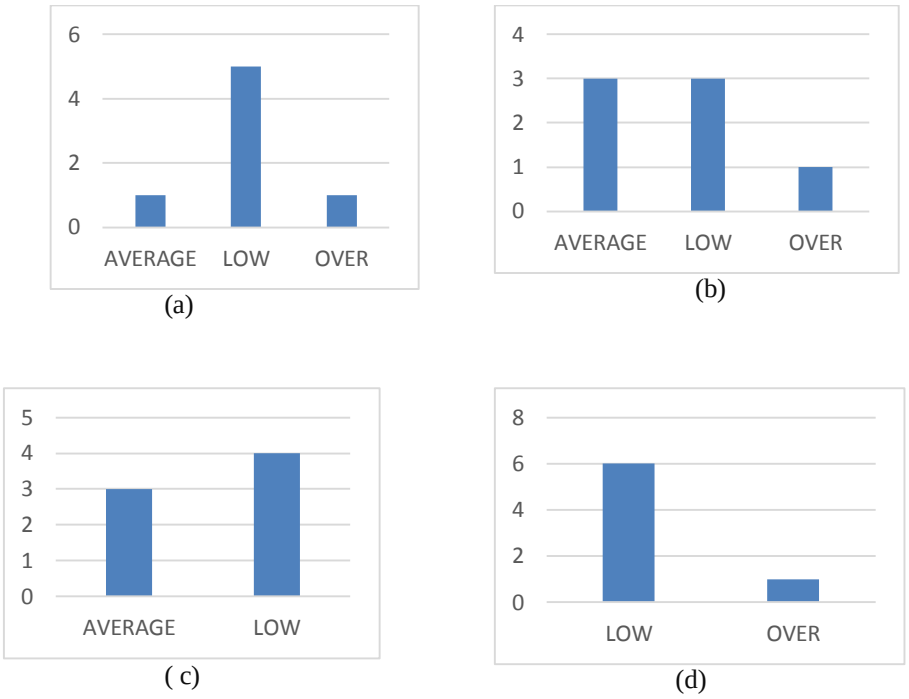


Figure 9: Analysis of the four resultant attributes

The statistics of the Naïve Bayes are shown in Table 1. The confusion Matrix is $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix}$. The above matrix is classified as a, b, c i.e. a= OVER, b= LOW, c = AVERAGE respectively.

V. CONCLUSION

For the people whoever suffering with the abnormal health issues the healthcare services are the necessary and they are treated as a most challenging aspects. Data mining plays a vital role over the health care industries, particularly in predicting several kinds of diseases.While predicting the diseases medical diagnosis is extensively used. Asa conclusion data mining is need not to be there for resolving issues in the healthcare data sets. We used to develop the hybrid model as a supportive to resolve the noticed issues. Firstly we used to record the daily activities of the patient's for a specific period of time at the three regular intervals. Later on, we applied a FP-growth to get association rules from the log file. Finally we have applied k-means clustering for an input together with the Bayesian network model to predict the health condition of a patient and advices the precautions accordingly.For the better enhancement in predicting along with a real time sensor meter data with the utilization of the hybrid models is as our future work.

REFERENCES

[1] Aldrich, F.K., 2003. Smart Homes: Past, Present and Future. Inside the Smart Home. In: Harper, R. (ed.). Springer-Verlag, London, pp. 17–39.

[2] Hargreaves, T., Hauxwell-Baldwin, R., Wilson, C., Coleman, M., Kane, T., Stankovic, L., Stankovic, V., Liao, J., Hassan, T., Firth, S., 2015. Smart Homes, Control and Energy Management: How Do Smart Home Technologies Influence Control over Energy Use and Domestic Life? European Council for an Energy Efficient Economy (ECEEE) Summer Study on Buildings, Hyeres, France.

[3] H. Zheng, H. Wang, and N. Black, "Human activity detection in smart home environment with self-adaptive neural networks," in Proc. of IEEE International Conference on Networking, Sensing and Control, pp. 1505-1510, 2008.

[4] E.M. Tapia, S.S. Intille, and K. Larson, "Activity recognition in the home using simple and ubiquitous sensors," Proceeding of Pervasive, vol. LNCS 3001, Springer Berlin Heidelberg, pp.158–175, 2005.

[5] Rashidi, P.; Cook, D.J. Transferring Learned Activities in Smart Environments, In Proceedings of the International Conference on Intelligent Environments, Barcelona, Spain, 20–21 July 2009.

[6] Riabov, A.; Liu, Z.; Wolf, L.; Yu, S.; Zhang, L. Clustering Algorithms for Content-Based Publication-Subscription Systems. In Proceedings of the 22nd International Conference on Distributed Computing Systems, Vienna, Austria, 2–5 July 2002; pp. 133–142.

[7] Li, C.; Biswas, G. Unsupervised Learning with Mixed Numeric and Nominal Data. IEEE Trans. Knowl. Data Eng. 2002, 4, 673–690.

[8] Alam, M.R.; Reaz, M.B.I.; Mohd Ali, M.A. SPEED: An Inhabitant Activity Prediction Algorithm for Smart Homes. IEEE Trans. Syst. Man Cybern. Part A Syst. Hum. 2011, 4, 985–990.

[9] R. Veloso, F. Portela, M. F. Santos, Á. Silva, F. Rua, A. Abelha, and J. Machado, "A Clustering Approach for Predicting Readmissions in Intensive Medicine," Procedia Technol., vol. 16, pp. 1307–1316, 2014.

[10] N. Sharma and H. Om, "Data mining models for predicting oral cancer survivability," Netw. Model. Anal. Heal. Informatics Bioinforma., vol. 2, no. 4, pp. 285–295, 2013.

[11] K.-J. Wang, B. Makond, and K.-M. Wang, "An improved survivability prognosis of breast cancer by using sampling and feature selection technique to solve imbalanced patient classification data," BMC Med. Inform. Decis. Mak., vol. 13, p. 124, 2013.

[12] H. M. Zolbanin, D. Delen, and A. Hassan Zadeh, "Predicting overall survivability in comorbidity of cancers: A data mining approach," Decis. Support Syst., vol. 74, pp. 150–161, 2015.

Author's Profile:

Ms. B. ArunaKumari, currently working as Assistant Professor at AITS, Rajampeta. She received her master's degree in Computer Science and Engineering from JNTUA pulivendula in the year 2015. She received her Bachelor's degree in computer science and engineering from CBIT, Proddutur affiliated to JNTUA University in 2012. Her areas of interest include Internet of Things and wireless networks

Ms. K.ROHINI, currently working as Assistant Professor at AITS, Rajampeta. She received her master's degree in Computer Science and Engineering from JNTUA University in the year 2015. She received her Bachelor's degree in Information Technology from RGMCET, Nandyal affiliated to JNTUA University in 2013. Her areas of interest include Internet of Things and wireless networks.