International Journal of Future Innovative Science and Engineering Research (IJFISER)
Volume – 3, ISSUE – 3 ISSN (Online):2454- 1966

Research Manuscript Title

A SURVEY ON MINING SMARTHOME BIG DATA FOR HEALTHCARE APPLICATIONS BY PREDICTING HUMAN ACTIVITY PATTERNS

J.Meena
PG Scholar
Computer Science and Engineering Department,
Vivekanandha College of Engineering for Women,
Namakkal, India.
Josmeena24@gmail.com

G.Rajasekaran
Assistant Professor
Computer Science and Engineering Department,
Vivekanandha College of Engineering for Women,
Namakkal, India.
Sekar4raja@gmail.com

Sep – 2017

www.istpublications.com

International Journal of Future Innovative Science and Engineering Research (IJFISER)
Volume – 3, ISSUE – 3 ISSN (Online):2454- 1966

A SURVEY ON MINING SMARTHOME BIG DATA FOR HEALTHCARE APPLICATIONS BY PREDICTING HUMAN ACTIVITY PATTERNS

J.Meena
PG Scholar
Computer Science and Engineering Department,
Vivekanandha College of Engineering for Women,
Namakkal, India.
Josmeena24@gmail.com

G.Rajasekaran
Assistant Professor
Computer Science and Engineering Department,
Vivekanandha College of Engineering for Women,
Namakkal, India.
Sekar4raja@gmail.com

Abstract: A 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 such transformation millions of homes are being set with smart devices (e.g. smart meters, sensors etc.) The existing system is a model of big data mining system that utilizes smart home big data as a means of learning and discovering human activity patterns for health care applications. This is used for frequent pattern mining, cluster analysis and prediction to measure and analyze energy usage changes sparked by occupants' behavior. Since people's habits are mostly identified by everyday routines, such as not preparing food or not using shower/bath. In this paper, we propose a human activity pattern mining model based on appliance usage variations in smart homes. The model which utilizes FP-growth for pattern recognition and Fuzzy C means clustering algorithms is capable of identifying appliance-to-appliance and appliance-to-time associations through automatic mining of energy consumption data. This is not only important to determine activity routines, but also, when utilized by health care application, is capable of detecting sudden changes of human activities that require attention by a health provider. We apply ANN for activity prediction based on individual and multiple appliance usage. 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.

Index Term: *Smart meters, Sensor, FP-growth, Fuzzy C means clustering algorithms, Energy consumption.*

I. INTRODUCTION

The recognition of domestic activities is based on the inference by fusing information from heterogeneous sensors and resolving the uncertainty due to the stochastic nature of human behavior and imperfect detection equipment. This also applies to the energy sector, especially for smart grid environments. The smart meters provide data that allows the power network to be more sophisticated. One of the most important problems in this area is to predict the power consumption as accurately as possible. Another interesting task is to segment customers according to their consumption profile. This can be useful to provide them with appropriate tariffs, adjust the power supply or issue regulations in the event of a power failure.

The Smart Energy systems have special characteristics that can influence the suitability of this approach. The question is: can we improve the prognosis of electricity consumption by first grouping the data into

groups and then applying forecasting methods. An interesting aspect is also the exploration of such features of forecasting methods, which use data decomposition into similar groups, which is also the subject of our interest.

The aim of this thesis is to discuss the possibilities for the detection and prediction of user activities in the Smart Home environment. If we can develop an accurate activity detection method, it can be implemented in the Smart Home Control System. Based on the activity detected by this method, the Smart Home can automatically provide the user with the appropriate service. Activity detection, however, is difficult in the real world due to the variability and complexity of the user activities that influence the accuracy of the recognition processes. Many researchers presented in the next chapter have suggested the possibilities for the recognition of activities, but the accuracy has not met the expectations. The reason is that they concentrated only on a part and not on the entire work. Overall, the activity recognition process consists of two parts. First, the repeated patterns of many activities must be discovered and classified. Secondly, what action the pattern means should be decided. Most of the earlier papers only dealt with one of these two points. It is difficult to achieve optimal performance when a problem is not explored from the overall view. The signal is detected by the mobile telephone:

Once the signal is detected, it is used for frequent pattern mining, cluster analysis, and prediction to measure and analyze changes in energy consumption caused by the behavior of the occupants. Since the destinies of man can be identified above all in everyday life, we recognize abnormal activities on these discoveries, which point to the difficulties of people to provide themselves. We propose a mountain model for human activity patterns based on the use of Smart Homes. The model that uses FP growth for pattern recognition and Fuzzy C, clustering algorithms, can be appliance and appliance consumption data. This is not only important to determine activity routines, but also when used by a health application to detect sudden changes in human activities observed by a health care provider. We use ANN for activity prediction based on the individual and multiple use of devices. For healthcare applications that integrate certain activities based on historical data for patients.

II.LITERATURE REVIEW

The model which utilizes FP-growth for pattern recognition and Fuzzy C means clustering algorithms is capable of identifying appliance-to-appliance and appliance-to-time associations through automatic mining of energy consumption data.We review each in literature turn.

In [1] authors Jing Liao, Lina Stankovic and Vladimir Stankovic in the paper **"Detecting Household Activity Patterns from Smart Meter Data"** proposed that there is a tough addiction on electrical appliances for domestic routines, here we propose an Non-intrusive appliance load monitoring (NALM) algorithm for identify domestic activities for smart meter aggregate data. We differentiate two types of behavior: Type I behavior are those that can be acknowledged using only smart meter data and Type II behavior are recognized by combining smart meter data with basic environmental sensing (temperature and humidity). By analyzing the both behaviour, we start by disaggregating the entire power produced down to individual electrical appliances.Then, we build an indicative activity model to reason four domestic behavior using the Dempster-Shafer theory of proof. To validate Non-intrusive appliance load monitoring (NALM) algorithms, we collect the environmental data of real energy in house hold during the period of months.Here we conclude that it is probable to notice four tested domestic daily behavior with high accuracy based on the aggregate energy usage.

In [2]authors Abdulsalam Yassine, Ali Asghar Nazari Shirehjini, And Shervin Shirmohammadi in the paper "**Smart Meters Big Data: Game Theoretic Model for Fair Data Sharing in Deregulated Smart Grids"** propose a game theoretic mechanism that balances between valuable uses of data and individuals' privacy in deregulated smart grids. Smart meters big data has large societal

importance that can be achieved by mechanisms that balance between consumers privacy and the benefit of sharing data. we presented a mechanism for sharing power consumption data in deregulated smart grids and also the concept of Activities in Daily Living as a means of data categorization and to help the data aggregator and the consumers to identify privacy risk values we assume that participants increase their cost linearly with the privacy risk value or with the anonymization level. These assumptions may require a more detailed modeling and user based studies to reflect reality. The concepts of "smart grid" and "smart meter " detected a significant gap between the knowledge of consumers and that they should have to take advantage of the new smart energy initiatives.

In [3]authors Qin Ni, Ana Belén García Hernando and Iván Pau de la Cruz in the paper "**The Elderly's Independent Living in Smart Homes:A Characterization of Activities and Sensing Infrastructure Survey to Facilitate Services Development**" propose a classification of the main activities considered in smart home scenarios which are targeted to older people's independent living, as well as their characterization and formalized context representation. Smart Home (SH) has gained a lot of attention for the terms of enhanced quality of life within the home. SH technologies may help to enhance quality of life, make longer independent living and reduce caregivers' necessary time and healthcare costs in general, without losing the safety that a continuous and unobtrusive monitoring provides. This is what is sometimes known as Ambient Assisted Living (AAL). AAL technologies and tools in terms of smart homes,assistive robotics, e-textile and wearable sensors, also exploring healthcare applications that focus on activity recognition algorithms and context modeling.They also use machine learning algorithms and statistical techniques to infer relevant information from the correspondence between the accelerometer and the electro myo gram (EMG)sensors measurements.The UML models show advantages in capturing information about the static structure and dynamic behaviour of a system, but have limitations in providing precise semantics and supporting reasoning on human behaviours.

In [4]authors Mohammad Arif Ul Alam, Nirmalya Roy, Michelle Petruska and Andrea Zemp in the paper **"Smart-Energy Group Anomaly Based Behavioral Abnormality Detection"** propose a data analytic approach that helps detect energy usage anomalies corresponding to the behavioral abnormality of the residents .It relies on detecting everyday appliances usage from smart meter and smart plug data traces in regular activity days and then learning the unique time segment group of each appliance's energy consumption. Nonintrusive load monitoring (NILM) research domain has been confined into the detection of types of appliances people have and their behavioral patterns through energy profiling and disaggregation.We apply other alternative methods such as GLDA, GMM and the K-Nearest Neighbor (KNN) algorithms that have been used successfully for collective anomaly detection for long time. it is possible to determine when prevention is necessary for an individual at risk.

In [5]authors M. Shamim Hossain in this paper **"Cloud-Supported Cyber–Physical Localization Framework for Patients Monitoring"** has drawn a great deal of interest from academia and industry. CCPSs facilitate the seamless integration of devices in the physical world (e.g., sensors, cameras, microphones,speakers, and GPS devices) with cyberspace. This enables a collection of emerging applications or systems such as patient or health monitoring, which require patient locations to be tracked. These systems integrate a large number of physical devices such as sensors with localization technologies (e.g., GPS and wireless local area networks) to generate, sense, analyze, and share huge quantities of medical and user-location data for complex processing. It presented an iPhonebased localization system for healthcare applications, whereby Wi-Fi signals and ZigBee RFID are used for localization. In their approach, the iPhone's position is first identified from the

received Wi-Fi signal using a fingerprinting algorithm. Users' mobility can be examined from mobile-phone location sensing mechanisms and cloud computing technologies for reliable localization. It introduced range-free localization algorithms for mobile clouds. It has benefits and limitations with regard to accuracy, ubiquity, and flexibility. A hybrid patient tracking localization technique is proposed to overcome this issue.

In [6]authors Kaustav Basu, Vincent Debusschere, Seddik Bacha in this paper "**Appliance Usage Prediction Using a Time Series Based Classification Approach"** proposed a model tries to formalize such an approach using a time-series based multi-label classifier which takes into account correlation between different appliances among other factors. The objective is to construct a model able to predict the appliance usage in housing which helps the system to arrange energy production and consumption and to decide which appliance will be used at each hour. The proposed model uses an iterative learning approach and tries to take into account all the possible information based on consumption data, time of the event and meteorological informations. The approaches range from using methodologies such as similar day, expert knowledge and linear and non linear learning algorithms.It gives details of implementation of neural networks in the domain of energy load forecasting and [lO] proposed a SVM model to predict daily load demand for a month. Another is to modify an existing single-label algorithm directly for the purpose of multi-label classification.

In [7]authors Kaustav Basua, Lamis Hawaraha, Nicoleta Arghira, Hussein Joumaaa, Stephane Ploixa in this paper **"A prediction system for home appliance usage"** has a novel about Power management in homes and offices requires appliance usage prediction when the future user requests are not available .This improve learning algorithms with expert knowledge and proposes a general model using a knowledge driven approach to predict if a particular appliance will start during a given hour or not .It is tested over the IRISE data and using different machine learning algorithms. The approaches range using methodologies such as similar day, expert knowledge and linear and nonlinear learning algorithms. The tested classifier algorithms were also introduced. These classifiers are used in order to choose the best one (which gives best accuracy) for each service in a house.

In [8]authors Yi-Cheng Chen, Hsiu-Chieh Hung, Bing-Yang Chiang, Sheng-Yuan Peng, Peng-Jun Chen in this paper **"Incrementally Mining Usage Correlations among Appliances in Smart Homes"** propose a great advent of sensor technology, residents can collect household appliance usage data easily. However, in general, usage data are generated progressively; visualizing how appliances are used from huge amount of data is challenging. Thus, an algorithm is needed to incrementally discover appliance usage patterns. Prior studies on usage pattern discovery are mainly focused on mining patterns while ignoring the incremental maintenance of mined results. In this paper, a novel method, Dynamic Correlation Miner (DCMiner), is developed to incrementally capture and maintain the usage correlations among appliances in a smart home environment. Furthermore, several optimization techniques are proposed to effectively reduce the search space.Experimental results indicate that the proposed method is efficient in execution time and possesses great scalability.Subsequent application of DCMiner on a real dataset also demonstrates its practicability.

In [9]authors Jiawei Han, Jian Pei, Yiwen Yin in this paper "**Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach**" propose a novel frequent-pattern tree (FP-tree) structure, a large database is compressed into a condensed, smaller data structure, FP-tree which avoids costly, repeated database scans. The Apriori heuristic achieves good performance gained by (possibly significantly) reducing the size of candidate sets. However, in such situations with a large number of frequent patterns, or quite minimum support thresholds, an Apriori-like

algorithm may suffer from the following two nontrivial costs: It is costly to handle a huge number of candidate sets. It is boring to repeatedly scan the database and check a large set of candidates by pattern matching, which is particularly true for mining long patterns . The benefits of FP-growth over Apriori becomes clear when the dataset contains an abundant number of mixtures of short and long frequent patterns.

In [10]authors Shailendra Singh, Abdulsalam Yassine, Shervin Shirmohammadi in this paper **"Incremental Mining of Frequent Power Consumption Patterns from Smart Meters Big Data**" propose incremental mining of frequent power consumption patterns from smart meters big data. The benefits of pattern growth strategy and mine in quantum of 24 hour period, i.e. frequent patterns are extracted from data comprising of appliance usage tuples for 24 hours period, in a progressive manner. Apriori algorithm with candidate generation approach suffers from two problems, it generates a large number of candidate sets and repeatedly searches through the entire database to find support for an item set. Similarly,it use incremental sequential mining technique to discover correlation patterns among appliances, and proposes a new algorithm offering reduction in memory with improved performance .It reduces the memory requirement on the cost of marginal increase in processing time, whereas former approach reduces processing time but requires more memory.

III.ANALYSIS

This section presents the study on human activity pattern mining model based on appliance usage variations in smart homes.

Table 1: COMPARITIVE STUDY ON MONITORING HUMAN ACTIVITY PATTERN

SI. NO	TITLE	ALGORITHMS USED	MERITS	DEMERITS
1.	Detecting Household Activity Patterns from Smart Meter Data	Non-intrusive appliance load monitoring (NALM)	Efficient energy management is improved	Difficult to change their lifestyles in order to save money
		Activity Recognition Algorithm		
		Inference Algorithm		
2.	Smart Meters Big Data: Game Theoretic Model for Fair Data Sharing in Deregulated Smart Grids	Game Theoretic Mechanism	Enormous economical, societal, and environmental Benefit.	Loss of privacy

3.	The Elderly's Independent Living in Smart Homes: A Characterization of Activities and Sensing Infrastructure Survey to Facilitate Services Development	Activity detection algorithms Machine learning algorithms Activity Recognition Algorithms Unsupervised learning algorithm. Threshold detection algorithm	Multiple-users activity monitoring and information about users is accurate.	Reliability constrains
4.	Smart-Energy Group Anomaly Based Behavioral Abnormality Detection	Hierarchical group anomaly detection algorithm Mixture of Gaussian Mixture Model (MGMM) algorithm Gaussian Latent Dirichlet Allocation (GLDA) algorithm K-Nearest Neighbor (KNN) algorithms	Cost and Privacy concerns are limited Used in large scale without additional Sensor	Real World testing has not to be done.
5.	Cloud-Supported Cyber–Physical Localization Framework for Patients Monitoring	Fingerprinting algorithm Range-free localization algorithms Maximization algorithm	Accuracy, ubiquity, and flexibility	Time processing, storage infrastructure is high.
6.	Appliance Usage Prediction Using a Time Series Based Classification Approach	Linear and non linear learning algorithms Single-label algorithm multi-label classification algorithms Decision tree algorithms Rakel algorithm	Data and time consumption is low.	Energy management and learned model should be predicted for better performance.
7.	A prediction system for home appliance usage	Neural network algorithm Nearest neighbour algorithm Decision tree algorithm	Knowledge representation is used to decreases the time .	Global Architecture is not suitable for all home Applications.

8.	Incrementally Mining Usage Correlations among Appliances in Smart Homes	Dynamic Correlation Miner (DCMiner) algorithm Apriori Algorithm	It reduce the search space. More efficient and increases mining correlation patterns.	Computational difficulty.
9.	Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach	Apriori algorithm TreeProjection algorithm. H-mine algorithm CLOSET algorithm PrefixSpan algorithm	Fault tolerant-Pattern growth method used to avoid costly candidate generation.	Issues arises in Pattern growth method
10.	Mining Energy Consumption Behavior Patterns for Households in Smart Grid	k-means algorithms Apriori algorithm Dynamic programming algorithm	Energy Efficiency is high	Very costly and not feasible.

IV.POSSIBLE SOLUTION

We propose a mining model for human activity patterns based on the use of Smart Homes. The model that uses FP growth for pattern recognition and Fuzzy C, clustering algorithms can be appliance and appliance consumption data. This is not only important to determine activity routines, but also, when used by a health care application, to detect sudden changes in human activities observed by a health care provider. We use a ANN for activity predictions based on individual and multiple use of devices. For healthcare applications that integrate certain activities on the basis of historical data for patients. For higher system accuracy, the prediction model integrates the chances of appliance-to-appliance and appliance-to-time associations, thereby identifying activities that are more specific in certain patterns.

V. CONCLUSION

In fuzzy inference models provide an proficient way to get over a patient performance quantitatively. The number of fuzzy rules acquired by human experts was always incomplete and inconsistent. This Study describes fuzzy clustering and combines techniques of fuzzy logic and neural network approach with sufficient training data sets in the determination of the Fuzzy rules from a fuzzy rulebase system of the student model. The Fuzzy Clustering and neuro-fuzzy system approach Problem of incompleteness in the decision by human experts. Completed Fuzzy rule base is formed by applying all possible fuzzy rules from the given Fuzzy variables and term sets. By training the neuro-fuzzy system with selected Samples that are safe should be the suggested approaches Decisions that could not be determined before. Although the fuzzy clustering and neuro-fuzzy system approach have successfully delivered the desired results any given input, this rule base is inconsistent in

that some of their rules have illogical conditions. Therefore, the increments are the training Records and several experiments were performed to model the models ANFIS and, finally, there have been acceptable classification results.

REFERENCES

[1] J. Liao, L. Stankovic, and V. Stankovic, ``Detecting household activity patterns from smart meter data," in Proc. Int. Conf. Intell. Environ. (IE),vol. 6. Jul. 2014, pp. 71_78.

[2] A. Yassine, A. A. N. Shirehjini, and S. Shirmohammadi, ``Smart meters big data: Game theoretic model for fair data sharing in deregulated smart grids," IEEE Access, vol. 3, pp. 2743_2754, 2015.

[3] Q. Ni, A. B. G. Hernando, and I. P. de la Cruz, ``The Elderly's independent living in smart homes: A characterization of activities and sensing infrastructure survey to facilitate services development," Sensors, vol. 15, no. 5, pp. 11312_11362, 2015. [Online]. Available: http://www.mdpi.com/1424-8220/15/5/11312.

[4] M. Ul Alam, N. Roy, M. Petruska, and A. Zemp, ``Smart-energy group anomaly based behavioral abnormality detection," in Proc. IEEE Wireless Health (WH), Oct. 2016, pp. 1_8.

[5] M. S. Hossain, "Cloud-supported cyber-physical localization framework for patients monitoring," IEEE Systems Journal, vol. 11, no. 1, pp. 118–127, March 2017.

[6] K. Basu, V. Debusschere, and S. Bacha, ``Appliance usage prediction using a time series based classication approach," in Proc. IEEE 38th Annu. Conf. Ind. Electron. Soc. (IECON), May 2012, pp. 1217_1222.

[7] K. Basu, L. Hawarah, N. Arghira, H. Joumaa, and S. Ploix,``A prediction system for home appliance usage," Energy Buildings, vol. 67, pp. 668_679, Sep. 2013. [Online]. Available:

[8] Y. C. Chen, H. C. Hung, B. Y. Chiang, S. Y. Peng, and P. J. Chen, "Incrementally mining usage correlations among appliances in smart homes," in Network-Based Information Systems (NBiS), 2015 18th International Conference on, 9 2015, pp. 273–279.

[9] J. Han, J. Pei, Y. Yin, and R. Mao, ``Mining frequent patterns without candidate generation: A frequent-pattern tree approach," Data Mining Knowl. Discovery, vol. 8, no. 1, pp. 53_87, 2004.

[10] S. Singh and A. Yassine, ``Mining energy consumption behavior patterns for households in smart grid," IEEE Trans. Emerg. Topics Comput., to be published.

[11] J. Clement, J. Ploennigs, and K. Kabitzsch, ``Detecting activities of daily living with smart meters," in *Advance Technology and Societal Change*.Heidelberg, Germany: Springer, 2014, pp. 143_160.

[12] C. Chalmers,W. Hurst, M. Mackay, and P. Fergus, ``Smart meter pro_ling for health applications," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*,Jul. 2015, pp.17.

[13] M. S. Hossain, "A patient's state recognition system for health care using speech and facial expression", *J. Med. Syst.*, vol. 40, no. 12,pp. 272:1_272:8, Dec. 2016.

[14] N. C. Truong, J. McInerney, L. Tran-Thanh, E. Costanza, and S. D. Ramchurn, ``Forecasting multi-appliance usage for smart home energy management," in *Proc. 23rd Int. Joint Conf. Artif. Intell.*, 2013, pp. 2908_2914.

[15] D. Schweizer, M. Zehnder, H. Wache, H. F. Witschel, Using consumer behavior data to reduce energy consumption in smart homes, 2015 IEEE 14th International Conference on Machine Learning and Applications(ICMLA), Miami, FL, USA, pp 1123 - 1129 London, UK: IEEE,2015.

[16] J. Kelly, W. Knottenbelt, The UK-DALE dataset, domestic appliance level electricity demand and whole-house demand from five UK homes.,Sci. Data 2:150007 doi: 10.1038/sdata.2015.7. London, UK: Scientific Data, 2015.