International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
Impact Factor: 5.22 (SJIF-2017), e-ISSN: 2455-2585
Volume 5, Issue 11, November-2019

Association Rules for Electrical Activity Detection in Smart Home to Reduce Electricity Wastage.

Wakiss Saiyed[1], Shilpa Patel[2]

[1]Computer Department, KITRC Kalol,
[2] Computer Department, KITRC Kalol,

Abstract*— Currently the use of big-data is on a large scale and is increasing day-by-day. So to handle the massive data is very important. In our research we have used Association Rules to generate rules from the best patterns which are generated by the Feature Selection. We have used Apriori Algorithm to generated Association Rules for finding best patterns to reduce electricity wastage. As it is a big-data means lots of data in high volume and also we have to provide high dimensionality as we are using feature selection so we have to be very carefully manipulate this all operations. In our research feature selection selects the features which are more dominating and it skips which are less dominating to get the better throughput using Best-Fit Linear Regression. As a result we can let the manufacturers know about the appliances usage at which time they are in use and in parallel or at at which time they are not in used. Our work is to find out the better patterns for electricity consumption in smart homes.

***Keywords*— Big-Data, Hadoop, Feature Selection, Best Fit Linear Regression, Apriori**

I. Introduction

Big-Data are such type of data that are structured semi-structured and un-structed. The data are in large volumes and is collected by particular organization to be get mined for further usefull approaches. Big data are generated on large scale due to which they are mostly generated by the social media or websites.

Increasing energy efficiency & reducing the unnecessary electrical wastage is the main aim behind the smart home approach. To reduce electrical wastage is achieved by applying the smart meters to each individual electrical appliances. Smart meters collect all the required data with which we will able to know at what time the appliance is in active state or in an in-active state. The opportunity to collect real-time consumption data prepares us to contemplate real-time feedback to inform the residents about their usage of energy [1].

The main purpose is to obtain the best efficiency for smart homes by reducing the electrical wastage and that is achieved by identifying the co-relations between the electrical appliances within the appropriate interval of time to achieve best time complexities. Cost saving is also the main purpose behind the smart home as the electrical wastage is reduce which leads to the savings in cost, acs explained by harper[2]. According to previous study by digitalSTROM energy savings can only be achieved if proper care regarding residents comfort is taken[3].

Association rules can find the existence of things that may exist or contact from the data behind. Which could be applied to the association study of user behavior in the smart home. The most classical association rule algorithm is Apriori algorithm. Apriori algorithm is an iterative manner based mining method by searching frequent item set.

According to IEEE 2016 “Temporal association rules for electrical activity detection in residential smart home” their main aim is to provide temporal sequential association rules in a novel way, based on machine learning techniques, to learn time windows where a rule’s take place to exploit historical data and their statistical properties. There are variances for the electrical activity detection for example when cooking activity is running then at that time it is possible that when electric stove is on then dishwasher is off or when oven is on then electric stove may off. Similarly like this we have to find out the best patterns from the association rules.

In this paper we have applied Feature Selection & Best Fit Linear Regressing with Apriori Algorithm and that is discussed in our next section. This paper is organized in total VI sections. Section II contains related work, Section III contains the work detail for which we have got idea for improvement. Section IV contains our proposed methodology, Section V contains the result analysis and last Section VI has conclusion and future work.

II. Related work

In the paper: “Temporal Association Rules For Electrical Activity Detection in Residential Homes” [4] [5] by “Hông-Ân Cao, Tri Kurniawan Wijaya, Karl Aberer And Nuno Nunes”, they have used data of 800 households for finding the best way to reduce the electric wastage.

They have used the following feature points for the testing and training from the dataset:
- Data Binarization
- Sequential Association Rules Mining

- Time Windows

1. Bivariate histograms (or heatmaps)
2. Tolerance regions

For their future research they have mentioned that they have derived time windows for temporal sequential association rules based on the co-occurrence of time intervals through machine learning techniques. Their novel method uses the statistical properties of the data to identify time windows without having to perform an additional or external search. The sequential events that can be seen as a bivariate histogram (or heatmap), which can be adjusted to guarantee that events arise in a significant enough proportion by applying a support threshold for the co-occurrence matrix. They have used Threshold to eliminate the noise from the variation of the time intervals and serves as the support filtering in the APRIORI algorithm[7] while computing the frequent itemsets each zone have strong co-occurrence can be approximated by a trivariate Gaussian distribution[11].

Daniel Schweizer, Michael Zehnder, Holger Wache, HansFriedrich Witschel, Danilo Zanatta and Miguel Rodriguez in their research 'Using consumer behavior data to reduce energy consumption in smart homes' [6] they have discusses how usage patterns of residents can be learned efficiently in smart home to achieve energy savings[8]. They have used a frequent sequential pattern mining algorithm for smart home event data. The performance of the proposed algorithm is compared to existing algorithms in terms of completeness or correctness of the results, run times as well as memory consumption. They also propose a recommender system based on the developed algorithm. This recommender provides recommendations to the users to reduce energy consumption. The recommender system was deployed to a set of test homes. They used this feedback to adjust the system parameters and make it more accurate during a second test phase. Digital STROM provided dataset that contains 33 homes with 3521 devices and over 4 million events[12]. This has produced 160 recommendations during the first phase and 120 durimg the second phase. The ratio of useful recommendations was near to 10%.

III. EXISTING WORK

For activities, the events used for the triggering and active usage of appliances in residential homes. The data that are recorded are in power data that do not contain information about when appliances are in active or running mode, instead of being off or in standby-mode. The measurements are converted to a binary form, where active and idle states are determined using GMMThresh [4], [5]. They want to get daily temporal association rules so they have considered that each day of collected data represents a basket, in the traditional market basket analysis [9][10]. The intervals during which the appliances are active represent the items in each basket. They have build a co-occurence matrix for window (heatmap) where each rule based on the time intervals. For time window (tolerance regions) The regions of interest for determining the time windows for the association rules are the temporal regions that occur the most often.

For their research work they have used the Dataport dataset, with data ranging from July 2012 until April 2015. They have used dataset in which they are applying time intervals for 1-minutews for each activity and then they are identifying the rules to get the best accuracy. Dataset contains 70 types of meters and more than 800 smart homes located mainly in Texas and in California. They have selected 16 households with large numbers of appliances. The measurements are binarized using GMMThresh [4], [5], which distinguishes when an appliance is active.

The first step is to read the dataset, then after they have performed the binarization of dataset. The results of the dataset are converted to binary form. Once the data are conerted to binary then data are to be mined for temporal mining. Co-occurrence matrix is generated from these dataset for further time window search. They have used GMM clustering for time window search from this a proper time window is obtained. Then after as per the time window the final rules are getting generated.

They have used Apriori algorithm for generating the association rules. With the help of association rules the patterns of the electrical appliances are identified and from these patterns the rules are generated by anayzing the patterns. For patterns analysing the threshold is ther where they have applied the time interval for 1 minute means during every minute they are identifying the status of the electrical appliances that at what time tha appliance are in used or in stand-by mode or in off mode.

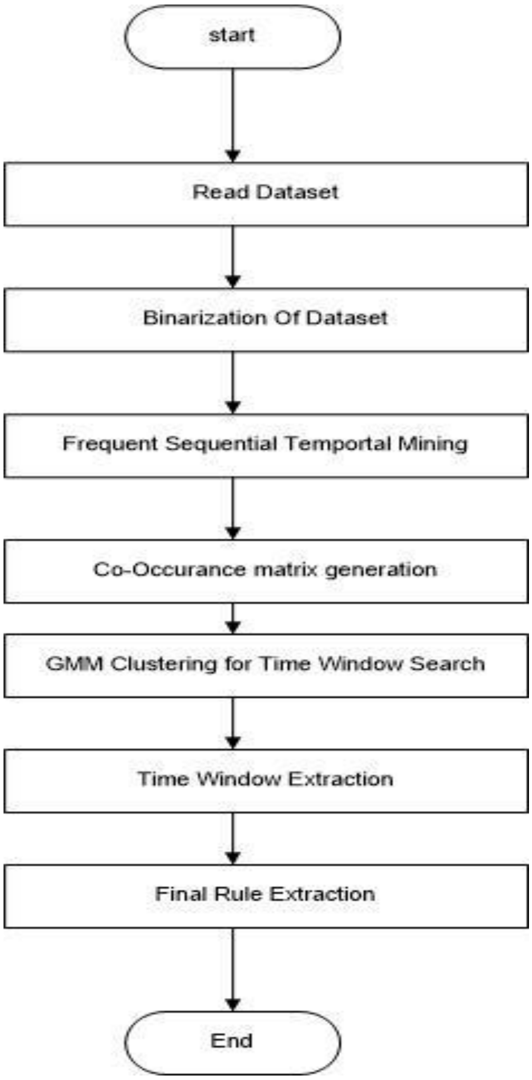

Figure 1
Current Flow Diagram

They summarize in Table I the total number of rules for all households. The number of association rules is the same for all methods depending on the chosen interestingness measure, as only the arrangements with enough support or confidence are selected, which guarantees enough data for the clustering.

TABLE I
RESULT FOR EXISTING SYSTEM

GMM	Covar.	Freq. Supp.	Interestingness	Min Score	Prob. Ellipse	Time Supp.	Total Nb. Rules	Total Nb. Temp Rules
DPGMM	diag	0.1	confidence	0.4	0.8	5	14214	55705
VBGMM	diag	0.1	confidence	0.4	0.8	5	14214	52840
GMM	diag	0.1	confidence	0.4	0.8	5	14214	68493
DPGMM	diag	0.1	support	0.4	0.8	5	8173	35048
VBGMM	diag	0.1	support	0.4	0.8	5	8173	33140
GMM	diag	0.1	support	0.4	0.8	5	8173	40079
DPGMM	full	0.1	confidence	0.4	0.8	5	14214	45783
VBGMM	full	0.1	confidence	0.4	0.8	5	14214	46513
GMM	full	0.1	confidence	0.4	0.8	5	14214	67769
DPGMM	full	0.1	support	0.4	0.8	5	8173	28876
VBGMM	full	0.1	support	0.4	0.8	5	8173	28674
GMM	full	0.1	support	0.4	0.8	5	8173	39859

IV. PROPOSED METHODOLOGY

Steps for Proposed Method

Step 1: Read Dataset

Step 2: Apply Best Fit Linear Regression

Step 3: Apply Feature Selection

Step 4: Binarization of Dataset

Step 5: Frequent Sequential Temportal Mining

Step 6: Co-Occuranc Matrix Generataion

Step 7: GMM Clustering for Time Window Search

Step 8: Time Window Extraction

Step 9: Final Rule Extraction

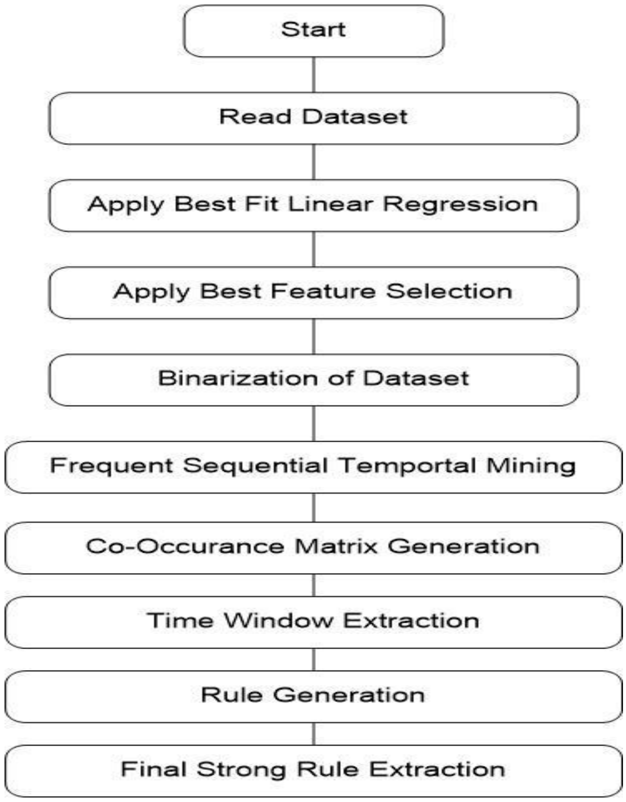

Figure 2
Proposed Flow Diagram

During the process we will first find out the best pattern by feature selection from the association rules and then we have applied the Best Fit Linear Regression for finding the parallel co-occurrence matrix to get the better throughput,
For getting the better result we have applied 15 minutes time interval due to which it is easy to manipulate all the operations and getting output properly.
Transactions contains multiple items which are running parallel from this transactions that contains multiple items we have finded Frequent Itemset.

V. RESULT ANALYSIS

We have got many Association Rules after processing below table contains some of the rules generated by our system.

ASSOCIATION RULE	SUPPORT	CONFIDENCE	LIFT
[WOE-Off,UTE-On] => [B1E-Off]	96.81%	100%	1.004 -> 0%
[FRE,EBE] => [B2E]	0.13%	100%	783.000 -> 78200%
[FGE-On,FRE-On] => [B2E-On]	45.59%	100%	1.004 -> 0%
[EQE,WOE] => [BME]	0.13%	100%	783.000 -> 78200%
[B1E-Off,HPE-Off] => [DNE-Off]	3.07%	86%	1.095 -> 9%
[WOE-On,FGE-Off] => [DNE-Off]	1.28%	83%	1.064 -> 6%
[UNE-Off,HPE-On] => [DNE-On]	0.13%	33%	1.563 -> 56%

VI. Conclusion and Future Work

Our main aim is to get the best association rules and to generate the best patterns as well as best co-relations between the devices. We are generating the rules with better interestingness because we have used apriori algorithm for generating the association rules and then after getting the best rules we are performing activity for final rule generation with the help of best fit linear regression method. We have also finded relations between various devices in terms of positive, negative or neutral. We got better result by using this approach where and still there is a better scope for future to improve time complexity and space complexity.

References

[1] K. Carrie Armel, A. Gupta, G. Shrimali, and A. Albert, "Is Disaggregation the Holy Grail of Energy Efficiency? The Case of Electricity," Energy Policy, vol. 52, pp. 213–234, Jan 2013.
[2] Harper, R."Inside the Smart Home" Springer-Verlag ISBN 978-185233-688-2, 2003 [3] Shuangxun Ma and Ruisheng Zhang, "PU-LP: A novel approach for positive and unlabeled learning by label propagation," *2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, Hong Kong, 2017, pp. 537-542. doi: 10.1109/ICMEW.2017.8026296
[3] Baumann, C. et al. "Smart energy case study" Proceedings ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings. BuildSys pp 36-38, Nov 2012. [5] Song, Yuqi&Gao, Min & Yu, Junliang& Li, Wentao& Yu, Lulan& Xiao, Xinyu. (2018). PUED: A Social Spammer Detection Method Based on PU Learning and Ensemble Learning: 13th International Conference, Collaborate Com 2017, Edinburgh, UK, December 11–13, 2017, Proceedings. 10.1007/978-3-030-00916-8_14
[4] H.-Â. Cao, T. K. Wijaya, and K. Aberer, "Estimating Human Interactions with Electrical Appliances for Activity-based Energy Savings Recommendations," in Proc. BuildSys '14. Memphis, TN, USA: ACM, Nov 2014, pp. 206–207.
[5] "Estimating Human Interactions With Electrical Appliances for Activity-based Energy Savings Recommendations," in Proc. BigData '16. Washington, DC, USA: IEEE, Dec 2016. [8] Taeho Jo, String Vector Based KNN for Text Categorization, ISBN 978-89-968650-9-4, ICACT2017 19 ~ 22, 2017
[6] King, N. "Smart home. A definition" Intertek Research and Testing Center, pp. 1-6, 2003
[7] R. Agrawal, T. Imielinski, and A. Swami, "Mining Association Rules Between Sets of Items in Large Databases," in Proc. SIGMOD '93. Washington, DC, USA: ACM, May 1993, pp. 207–216.
[8] Zehnder, M. "Energy saving in smart homes based on consumer behaviour data" Master's thesis, School of Business, University of Applied Sciences and Arts Nortwestern Switzerland (FHNW) Jan 2015.
[9] R. Agrawal, T. Imielinski, and A. Swami, "Mining Association Rules Between Sets of Items in Large Databases," in Proc. SIGMOD '93. Washington, DC, USA: ACM, May 1993, pp. 207–216.
[10] R. Agrawal and R. Srikant, "Mining Sequential Patterns: Generalizations and Performance Improvements," in Proc. ICDE '95. Taipei, Taiwan: IEEE, Mar 1995, pp. 3–14.
[11] C. B. Do, "The Multivariate Gaussian Distribution," 2008. [Online]. Available: http://cs229.stanford.edu/section/gaussians.pdf
[12] Dickmann, G. "digitalSTROM: A centralized PLC Topology for Home Automation and Energy Management" IEEE International Symposium on Power Line Communications and its Applications (ISPLC), pp 352357, April 2011.