A Design Approach for Efficient Mining Human Activity Patterns over Big Data

P. V. N. Rajeswari* and Kodi Roshan Sai Kumar

Department of CSE, Visvodaya Engineering College, Kavali – 524201, Andhra Pradesh, India; rajivrphd@gmail.com, roshansaikumar24@gmail.com

Abstract

Background/Objectives: In recent times, urban areas are turned to be a smart dewing spot. In smart cities, most things are done automatically using smart devices such as sensors and smart meters. These smart devices produce large volumes of fine-grained data and stored into servers. To make a study on how to train and detect human health behaviour patterns for healthcare applications from smart meter data. **Methods/Statistical Analysis**: To analyze and detect human activity patterns, we build an Efficient Mining technique using frequent pattern mining and clustering techniques. Here, we consider the UK-Dale dataset incorporates time series information of intensity utilization gathered somewhere in the range of 2012 and 2015. To training and detecting human pattern we build an Efficient Mining techniques with accuracy for short and long term predictions. **Applications/Improvements:** We also extend our work to analyze temporal energy consumption patterns at the appliance level, which is directly related to human activities.

Keywords: Behavioural Analysis, Big Data, Data Clustering, Frequent Pattern Mining, Smart Cities

1. Introduction

The ongoing overview demonstrates the large portion of the general population indicating enthusiasm to lives in urban territories¹. In reacting to the new needs and difficulties, urban communities are at present grasping monstrous advanced change with an end goal to help manageable urban networks, and give more advantageous condition^{2,3}. In such a change, a large number of homes are being furnished with shrewd gadgets, which create enormous volumes of fine-grained and indexical information that can be dissected to help social insurance services.

Progression of big data mining advances, which give methods for handling immense amount of information for noteworthy experiences, can help in seeing how individuals approach their life. For instance, checking the alterations of appliance utilization inside a smart home can be utilized implicitly decide the individual's prosperity depending on authentic information. Since individuals' propensities are for the most part distinguished by ordinary schedules, finding these schedules enables us to perceive odd exercises that may demonstrate individuals' troubles in taking consideration for themselves, for example, not making food or not utilizing shower/shower^{4.5}. The fundamental connection between appliance utilization inside the smart home and routine exercises can be utilized by health care applications to identify potential medical issues. This isn't just going to ease the burden on health issues, yet additionally giving 24 hour checking administration that naturally distinguishes typical and anomalous practices for autonomously living patients or those with self-constraining conditions (e.g. elderly and patients with cognitive impairments).

This study proposes the utilization of vital information from smart meters fixed at homes to uncover essential exercises of occupants. Our investigation accepts that there are methods set up to shield individuals' secrecy from being

*Author for correspondence

shared or assessed for unjustifiable uses in^{6.7}. The projected model discerns and explores readings from smart meters to acknowledge activities and changes in action.

Uncategorised power exploitation readings are explicitly connected with activities accomplished at home. To deal with the recently referenced issues, this article professes repetitive mining and expectation model to verify and analyze fundamental exploitation changes begun by occupants' conduct.

The data will be recursively mined in the quantum/ data slice of 24 hours via smart meters and the results are kept up transversely over subsequent mining exercises. Likewise we utilize the Bayesian framework, a probabilistic graphical model, to anticipate the usage of different appliances and household energy usage. The projected model is set up to do transitory figures extending from one hour to 24 hours and long haul expectation for a significant time span, weeks, months or seasons.

2. Research Method

It starts by verifying and setting up the data and later applying regular pattern mining for discovering appliance-to-appliance affiliations, i.e. appliances are cooperating. By then, it uses cluster analysis to choose appliance to-time associations. With these two methods, the structure can remove the case of appliance usage which is then used as a commitment to the Bayesian system. The yield of the framework is used by specific health applications relying upon the proposed use. Figure 1 depicts the proposed model.



Figure 1. Mining frequent patterns and activity predictions for health care applications in smart homes.

2.1 Data Preparation

The dataset exploited in this examination is an aggregation of smart meters information from five houses in the United Kingdom (UK)^{8.9}. This dataset incorporates 400 million raw records at time resolution of 6 seconds. In the primary phase of the cleaning procedure we created customized techniques to expel commotions from the information and set it up for mining. In the pre-processing step, the dataset is curbed to 20 million. In addition, we accomplished a processed dataset for fundamental analyzing of the model, having over 1.2 million records. Tables 1 and 2 exemplifies the subsequent strategy to mine source information including four appliances from one house. Smart meters time-series crude information, which is a high time-resolution information, is altered into a 1-minute resolution stack information; Likewise it is modified into a 30 minutes time-resolution source information, i.e. $24 \times 2 = 48$ readings for every day per appliance, while noticing start time and end time for every dynamic appliance.

2.2 Extracting Frequent Patterns for Human Activity

The objective is to find human activity patterns from smart meters information. For instance, exercises, for example, "Sitting in front of the TV, cooking, using computer, preparing food and cleaning dishes or clothes" are generally standard schedules. Our point is to identify the examples of these exercises with the goal that a health care application, that screens sudden changes in patients conduct (e.g. patients with intellectual weakness), can send convenient caution to health care providers. In seeking after such process, all appliances that are enlisted dynamic during the 30-minute time interim are incorporated into the source database for continuous pattern information mining. Extricating human movement patterns isn't just finding the individual appliance activity, yet in addition the appliance-to-appliance affiliations.

The central thought of the model projects pattern growth or FP-growth approach using depth-first divide and conquer methodology. Thus, we project another strategy that uses the benefits of pattern growth framework and extends it to achieve steady dynamic mining of consistent patterns by mining in 24 hours quantum progressively. The model's hidden idea relies on the^{10–12}, proposed setup, in which we guaranteed that each and every incessant patterns are stored in the database. In such way, recently discovered progressive patterns are compared with existing ones in the database and the support is upgraded or else the new pattern is incorporated the database. The extent of the database size is surged for all the progressive patterns when the mining method is done, which helps to analyze the data accurately. Algorithm 1 portrays progressive pattern mining process.

Algorithm 1 Incremental Frequent Pattern Mining.

Require: Transaction database (DB), Frequent pattern discovered database (FP_DB).

Ensure: Incremental discovery of frequent patterns, stored in frequent patterns discovered database (FP_DB).

1. For all Transaction data slice in quanta of 24 hours in database DB do {Data is processed in slices of 24 hour period}.

2. Determine database size for data slice/quantum

3. Mine Frequent patterns in using extended FP-growth approach.

4. For all Frequent Pattern *FP* in do.

- 5. Search a frequent pattern *FP* in
- 6. If Frequent Pattern found then.
- 7. Update frequent pattern in
- 8. Else
- 9. Add a new Frequent Pattern to
- 10. End if.
- 11. End for.

12. For all Frequent Patterns in Database increment Database Size by

13. End for.

1	ab	le 1	I. R	eady	to	mine	source	data	l

Date	ST	ET	Active Appliances		
2013-08-01	07:00	07:30	^{'2} 3 4 12'		
2013-08-01	07.30	08.00	<u>'3 4 12'</u>		
2013 08 01	07:50	08.30	³ 4 12		
2013-08-01	00:00	00.00	2412 (412)		
2013-08-01 08:30 09:00 412					
ST = Start Time, ET = End Time, 2 = Laptop,					
3 = Monitor, 4 = Speakers, 12 = Washing Machine					

2.3 Clustering Analysis: Incremental K-means

Finding appliance to-time affiliations is vital to wellbeing applications that screen patients' action designs on daily

basis. In this area, a clustering analysis mechanism is utilized to find appliance use time as for hour of day (00:00 - 23:59), time of day (morning, afternoon, evening, night), weekday, week as well as month of the year. Appliance-totime affiliations are fundamental data in the smart meter time arrangement information which incorporate adequately close time-stamps, when applicable appliance has been recorded as dynamic or operational. Utilizing this information we can aggregate a class or group of appliances that are in operation ubiquitously or overlapping^{13.}

Table 2. Clustering source database

Appliance	Hour of Day	Time of Day			
2	07:30 08:00	M E			
3	16:00 16:30	A E			
12	13:30 14:00 14:30	А			
2 = Laptop, 3 = Monitor, 12 = Washing Machine					
M = Morning, A = Afternoon, E = Evening					

We chose a 30 minute time range/cut, for cluster segmentation, which will adequately catch the affiliations while limiting the quantity of segments made; i.e., making greatest 48 clusters for a day, though other clustering bases, for example, time-of-day, weekday, week and months have regular segmentation.

In our model, incremental progressive clustering is received by combining existing and newfound groups of each progressive mining activity into the database. This incremental procedure is accomplished by ensuring all applicable cluster parameters, for example, SSE, Silhouette coefficient (width), information points and space from the centroid are recorded in the database. The algorithm that guarantees effectiveness and speed of such activity can be found. Figures 2 and 3 demonstrate the clustering investigation for time and day of the week for house 5 in the dataset. They showed two figures that illustrate the connections between the activity of appliance and the season of utilization. This data of appliance connections are exploited for predicting the actions inside houses.

2.4 Conjunctive Keyword Search Protocol

We coordinate the frequent patterns and appliance to-time relationship to find out about the utilization of numerous appliances and create the activity prediction model. The system uses a Bayesian network which is a coordinated non-cyclic graph, where nodes speak to irregular variables and edges show probabilistic conditions. One of the primary features of a Bayesian network is that it incorporates the idea of causality. Notwithstanding the structure, a Bayesian network display gives a conservative method for speaking to the joint probabilistic distribution. At the end of the day, every node or variable is free of its non-relatives and joined by its neighbourhood restrictive probability distributions as a node probability table, which encourages the calculation of the joint contingent probability dispersion for the model. An imperative benefit of the Bayesian system is the ability of reducing the missing data, learn associations. A Bayesian system is classified by the probabilistic distribution figured in below formula 1.

$$p(x_1, x_2, \dots, x_n) = \prod_{i=1}^n p(x_i \mid parents(x_i)) \quad (1)$$

As referenced over, our probabilistic prediction method is constructed by relying on consolidating probabilities for appliances to-time relationship to the extent per hour of the day (00:00 - 23:59), time of day (Morning, Afternoon, Evening, Night), weekday, week, month, season and appliances to-appliances affiliations. The topology of the consequent Bayesian system has only a solitary dimension of data verification nodes, joined by independent unequivocal probabilities, consolidating to one outcome node. Formula (2) displays the back probability or peripheral spread for the proposed predictionmethod^{14,15}.



Figure 2. Incremental clustering of appliances-to-time of day associations inside one of the five houses using 25% of the data.



Figure 3. Incremental clustering of appliances-to-time of week associations inside one of the five houses using 25% of the data.

3. Results and Analysis

For the evaluation of the proposed model, we accomplish our investigations using the dataset UK-Dale together with the built dataset to inspect middle and last results. The (UK-Dale) dataset consolidates time series data of power exploitation which is collected in the range of 2012 and 2015.

The fundamental stage in understanding human actions is by removing relationship of appliances use. Figures 4 and 5 exhibits the appliances to-time affiliations found for the time and weekday individually for house 2. We can see that somewhere in the range of 2:30 and 5:00 PM on TV, Toaster, Living room Lights are utilized together in this house with the most astounding fixation amid the end of the week. Likewise, the clothes washer and Laptop are all the while utilized somewhere in the range of 8:30 and 10 am. The clothes washer is utilized all weekdays, where the Laptop isn't utilized on the ends of the week. Considering these actualities we can see the fluctuating impact of time and days on the utilization of appliances. The outcome is for 3 houses and it depends on handling 25% of the dataset. One can without much of a stretch see from appliance affiliations that tenants of house 1 get a kick out of the chance to unwind while getting ready food.

The prediction model uses appliance-to-appliance and appliance-time relationship to anticipate different synchronized appliances. Figure 6 demonstrates the probability distribution of appliances for house 2 over time of the day and weekday.



Figure 4. House 3: Appliance-time associations hour of the day.



Figure 5. House 3: Inter appliances associations day of the week.



Figure 6. Detected appliances with probability distribution over 24 hours period.

4. Conclusion

In this study, we presented a model for monitoring human activity patterns from low promising smart meters data. Occupant's actions and conduct follow a pattern that could be used in health applications to pursue the success of individuals living alone or those with self-compelling conditions. We presented incremental frequent mining and prediction method which relies on Bayesian system. In our present work, through experimentations, we found that the 24-hour term was perfect for data mining, yet we made the model to deal with any time quantum. From the results of examination, we have displayed the significance of the proposed model to perceive different appliances uses adequately and make a short and long term prediction at high precision.

5. References

- World urbanization prospect. United Nation.2017. https:// population.un.org/wup/
- Hossain MS. Cloud-supported cyber-physical localization framework for patients monitoring. IEEE Systems Journal. 2017; 11(1):118–27. https://doi.org/10.1109/ JSYST.2015.2470644
- Hossain MS, Muhammad G, Abdul W, Song B, Gupta B. Cloud-assisted secure video transmission and sharing framework for smart cities. Future Generation Computer Systems. 2018; 83:596–606. https://doi.org/10.1016/j. future.2017.03.029
- Liao J, Stankovic L, Stankovic V. Detecting household activity patterns from smart meter data. International Conference on Intelligent Environments. 2014; 6:71–8. https://doi.org/10.1109/IE.2014.18
- Yassine A, Shirehjini AAN, Shirmohammadi S. Smart meters big data: Game theoretic model for fair data sharing in deregulated smart grids. IEEE Access. 2015; 3:2743–54. https://doi.org/10.1109/ACCESS.2015.2504503
- Yassine A, Shirmohammadi S. Measuring users' privacy payoff using intelligent agents. https://doi.org/10.1109/ CIMSA.2009.5069940
- Yassine A, Shirmohammadi S. A business privacy model for virtual communities. International Journal of Aeb based Communities. 2009; 5(2):313–35. https://doi.org/10.1504/ IJWBC.2009.023971
- Chen YC, Hung HC, Chiang BY, Peng SY, Chen PJ. Incrementally mining usage correlations among appliances in smart homes. International Conference on Networkbased Information Systems; 2015. p. 273–9. https://doi. org/10.1109/NBiS.2015.43
- The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes. 2015. https://www.nature.com/articles/sdata20157
- Han J, Pei J, Yin Y. Mining frequent patterns without candidate generation. Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data. 2000; 29(2):1–12. https://doi.org/10.1145/342009.335372

- Han J, Pei J, Yin Y, Mao R. Mining frequent patterns without candidate generation: A frequent-pattern tree approach. Data Mining and Knowledge Discovery. 2004; 8(1):53–87. https://doi.org/10.1023/B:DAMI.0000005258.31418.83
- Chalmers C, Hurst W, Mackay M, Fergus P. Smart meter profiling for health applications. International Joint Conference on Neural Networks (IJCNN); 2015.
 p. 1–7. PMCid: PMC4646133 https://doi.org/10.1109/ IJCNN.2015.7280836
- 13. Han J, Kamber M, Pei J. Data Mining: Concepts and Techniques. The Morgan Kaufmann Series in Data Management Systems;

2012. p. 1-740. https://doi.org/10.1016/B978-0-12-381479-1.00001-0

- 14. Introduction to Bayesian Networks. 2018. https:// towardsdatascience.com/introduction-to-bayesian-networks-81031eeed94e
- 15. Pearl J. Probabilistic reasoning in intelligent systems: Networks of Plausible Inference. Morgan Kaufmann; 1988. p. 1–552. https://doi.org/10.1016/B978-0-08-051489-5.50007-2