Research Article **ISSN: 2395 – 5597**

International Journal of Engineering and Robot Technology

Journal home page: www.ijerobot.com

A TIME BASED MARKOV MODEL TO FORMULATE THE DYNAMICS OF CUSTOMERS ELECTRICITY CONSUMPTION BEHAVIOR

S. Jagadesh*[1] and A. Mahesh[1]

[1]*Department of Computer Science and Engineering, Anna University, Chennai, India.

ABSTRACT

In a competitive retail market, giant volumes of sensible meter knowledge give opportunities for load serving entities to boost their data of customers' electricity consumption behaviors via load identification. Rather than that specialize in the form of the load curves, this paper proposes a unique approach for a cluster of electricity consumption behavior dynamics, wherever dynamics" hash out with transitions and relations between consumption behaviors, or rather consumption levels, in adjacent periods. First, for every individual client, symbolic combination approximation is performed to scale back the dimensions of the information set, and time-based Markov model is applied to model the dynamic of electricity consumption, remodeling the massive knowledge set of load curves to many state transition matrixes. Second, a clustering technique by quick search and notice of density peaks (CFSFDP) is primarily doling out to get the everyday dynamics of consumption behavior, with the distinction between any 2 consumption patterns measured by the Kullback-Liebler distance, and to classify the purchasers into many clusters. To tackle the challenges of huge knowledge, the CFSFDP technique is integrated into a divide-and-conquer approach toward big data applications. A numerical case verifies the effectiveness of the proposed models and approaches.

KEYWORDS

Load profiling, Hadoop, Map Reduce, Big data, Markov model, Electricity consumption, Behavior dynamics, Distributed clustering and Demand response.

Author for Correspondence:

Jagadesh S,
Department of Computer Science and Engineering,
Anna University, Chennai, India.

Email: jagadesh477@gmail.com

INTRODUCTION

Countries round the world have set aggressive goals for the restructuring of noncompetitive grid towards liberalized markets particularly on the demand aspect. During a competitive retail market, load serving entities (LSEs) are going to be developed in nice numbers. Having a stronger understanding of electricity consumption patterns and realizing customized power managements area unit effective ways that to boost the aggressiveness of LSEs. Meanwhile, sensible grids are revolutionizing the electrical generation and consumption through a two-way flow of power and knowledge. As a vital info supply from the demand aspect, advanced metering infrastructure (AMI), has gained increasing quality worldwide; AMI permits LSEs to get electricity consumption information at high frequency, e.g., minutes to hours. Massive volumes of electricity consumption information reveal info of consumers which will doubtless be utilized by LSEs to manage their generation and demand resources with efficiency and supply customized service[1-3].

Load identification that refers to electricity consumption

Behaviors of consumers over a particular amount, e.g., one day, can facilitate LSEs perceive however electricity is really used for different customers and acquire the customers' load profiles or load patterns. Load identification plays a significant role within the Time of Use (ToU) tariff style[4], nodal or client scale load forecasting[5], demand response and energy potency targeting[6], and non-technical loss (NTL) detection[7]. The core of load identification is cluster which may be classified into 2 categories: direct cluster and indirect clustering[8]. Direct cluster implies that cluster ways are applied on to load information. Heretofore, there square measure an out sized number of cluster techniques that square measure wide studied, including k-means[9], fuzzy k-means[10], class-conscious clustering[11], self-organizing maps (SOM)[12], support vector cluster[13], topological space cluster[14], ant colony clustering[15] and etc. The performance of every cluster technique can be evaluated and quantified victimisation varied criteria, as well as the cluster dispersion indicator (CDI), the scatter index (SI), the Davies Bouldin index (DBI), and the mean index adequacy (MIA)[16]. The deluge of electricity consumption information with the wide spread and high-frequency assortment of sensible meters introduces nice challenges for information storage, communication and analysis. During this context, dimension reduction ways will be effectively applied to cut back the dimensions of the load information before clustering that is outlined as indirect cluster. Such clustering will be classified into 2 sub-categories, feature.

Extraction-based cluster and time series-based cluster

Feature extraction that transforms the info within the high dimensional space into area| an area} of fewer dimensions[17], is often accustomed scale back the dimensions of the input file. Principal component analysis (PCA)[18,19] may be aoftimes used linear dimension reduction technique. It tries to retain most of the covariance of the info options with the fewest artificial variables. Some nonlinear dimension reduction ways.

Including Sammon maps, curvilineal element analysis

(CCA)[20], and deep learning[21] have conjointly been applied to electricity consumption information. Moreover, as electricity consumption information square measure basically a statistic. A spread of mature analytical ways like separate Fourier remodel (DFT)[22,23].

Separate wave remodel (DWT)[24], symbolic combination approximation (SAX)[25], and therefore the hidden Markov model (HMM)[1] are mentioned within the literature. These ways square measure capable of reducing the dimensionality of your time series and of maintaining a number of the original character of the electrical consumption information.

BASIC METHODOLOGY

The projected methodology for the dynamic discovery of the electricity consumption is divided into six stages, as shown in Figure No.2. The primary stage conducts some load information preparations, as well as information improvement and cargo curve normalization. The second stage reduces the spatial property of the load profiles exploitation SAX. The third stage formulates the electricity consumption dynamics of every individual client utilizing time-based Andre Mark off model. The K-L distance is applied to live the distinction between any 2 Andre Mark off model to get the space matrix within the fourth stage. The fifth stage performs a changed CFSFDP clump algorithmic program to discover the standard dynamics of electricity consumption.

Data preparations including data cleaning is not the subject of this paper and will not be discussed. To make the load profiles comparable, the normalization process transforms the consumption data of arbitrary value {} 1 2, H x =xx Abbreviations and Acronyms.
X to the range of (0, 1), as shown in (1). Other Recommendations
It should be noted that the normalization is performed on a day after day rather than over entire periods. This strategy is chosen for at least three reasons. First, it can weaken the impact of anomalous days with critical peaks or bad data injections. Second, it can provide load shapes with little effect from daily or seasonal changes in the maximum values. Third, it can filter out the base load, which has little effect on demand response and reserve, in favor of the fluctuant part, which shows greater potential in demand response[26-28].

SAX for Load Curves

SAX is a powerful technique for the dimensional reduction and illustration of your time series knowledge with lower bounding of the Euclidean distance[29]. SAX discretizes numeric time series into symbolic strings by two steps: transforming the load data into a piecewise aggregate approximation (PAA) representation and then symbolizing the PAA representation into a discrete string. The basic idea of PAA is intuitive and easy, commutation the amplitude values falling within the same amount with their mean values, as shown in (2). If you are using Word, use either the Microsoft Equation Editor or the Math Type add-on (http://www.mathtype.com) for equations in your paper (Insert | Object | Create New | Microsoft Equation or Math Type Equation). "Float over text" should not be selected.

Distance Calculation

The dissimilarity/distance measurement is a fundamental problem in clustering. There exist many ways to compute the distances between two matrices, such as 1-norm distance and 2-norm distance (Euclidean distance). However, totally different from general matrices, a N N× state transition probability matrix essentially consists of N probability distributions, where each row (e.g., the ith row) corresponds to a probabilistic distribution of the state of the next period at the current state (e.g., the ith state). K-L distance is an effective way to quantify the dissimilarity between two probabilistic distributions[30]. Thus, for discrimination between two Markov model with the state transition matrices t Pi and t Pj, the K-L distance is defined as[2].

DISTRIBUTED ALGORITHM FOR LARGE DATA SETS

The electricity consumption knowledge sky rocketing for population-level customers, is difficult the storage, communication and analysis of the data. Although SAX and time-based Markov model have largely reduced the dimensionality of the load profiles, the centralized clustering technique is not effective for big data challenges. On one hand, the electricity consumption data are collected and distributed on completely different sites. The electricity consumption knowledge of consumers are collected and stored on different substations they belong to. It is costly and time consuming to transmit whole information from every distributed web site} to a central site. On the opposite hand, the analysis and clustering of large data sets gathered from each distributed site need a very large time and memory overhead. Once applying the CFSFDP, the un similarity matrix of all the customers should first be obtained, which accounts for most of the computation time. Both the time and space complexity of the CFSFDP are n fact, there exist many works on parallel clustering for big data applications[3,4] For these algorithms, the whole data set should reside on the same data center and then be distributed to different clients like map-and-reduce in Hadoop. It is not satisfied with the practical situation of electricity consumption data collecting and storing. Besides, some totally distributed cluster algorithms are projected to tackle these challenges by aggregating the information of local data and then sending to a central web site for central analysis. However, these algorithms don't contain the benefits as CFSFDP. Thus, this section is proposed to design a fully distributed instead of parallel clustering rule to ease the communication and computation burden further as retain the advantages of the CFSFDP by a divide-and-conquer framework. A. Framework Figure No.1 gives a divide-and-conquer framework for distributed clustering, where Li denotes the original data on the its distributed local site; Mi denotes the representative objects selected from the ith distributed local site; and R denotes the worldwide agglomeration results. Every object corresponds to a client delineated by transition probability matrixes. The proposed algorithm consists of three steps: Step 1: The SAX and time-based Markov model for individual customers are handled separately. Divide the big data set into k parts, each marked as Li. Note that the data on one distributed site can be additional divided to create the scale of the information sets on every website additional even. Step 2: An adaptive k-means method is performed for each individual part to obtain a certain number of cluster centers. Every cluster center will represent all the objects happiness to the present cluster with a small error. All these cluster centers of Li are selected as the representative objects Mi, which are defined as a local model.

Local Modeling-Adaptive k-means A set of clustering centers will be obtained by k-means, where the sum of the squared distances between each object is reduced. These centroids are often used as a "code book": every object are often described by the corresponding centroid with the least error. This is called vector quantization (VQ). We try to establish a local model by finding the "code book" that guarantees that the distortion of each object by VQ satisfies the threshold condition according to (11) where t Ckij denotes the kth centroid; and θ denotes the distortion threshold. Traditional k-means needs a given number of centers, which makes it difficult to guarantee that (11) holds. In this paper, an adaptive k-means is adopted to dynamically change the quantity of centers following an easy rule: if an object violates the threshold condition, 2-means (i.e. k-means for k=2) will be applied to partition this cluster further and add a new center to the "code book"[30]. The Figure No.2 hows the detailed procedures of the adaptive k means method. The distortion threshold θ varies depending on different needs. Smaller threshold corresponds to higher clustering accuracy and larger number of local representative objects, and vice versa. As a supplement of distortion threshold and another terminating condition of the iteration, Global Modeling-Modified CFSFDP The original CFSFDP algorithm considers the clustered objects equally. However, in a two-level clustering framework, the chosen representative models from completely different native sites would possibly represent "samples" of different populations. It would be reasonable to consider the representativeness of the local models in the centralized clustering. Thus, a modified CFSFDP method is proposed, which introduces a weight factor to differentiate the representativeness of the local models. Without loss of generality, the weight factor, Cj, is added to the local density calculation.

CASE STUDIES

Description of the information Set the information set utilized in this paper was provided by analysis Perspective, Ltd. and contains the electricity consumption of 6,445 customers (4, 511 residents, 391 industries, and 1533 unknown) over one and a 0.5 years (537 days) at a coarseness of half-hour[20]. The whole data set consists of total 3.46 million (6445 537×) daily load profiles. The bad load profiles are roughly identified by detecting the load profiles with missing values or all zeroes. Among these massive load data, we eliminated 6187 bad load profiles, which is a very small sample (approximately 0.18%) of the whole data set. B. Modeling Dynamics of Electricity Consumption for Each Customer According to the regular routine of electrical customers, we reasonably divide on a daily basis into four periods: amount one (00:00- 06:30, 22:00-24:00, long period), Period 2 (06:30-11:30, morning period), Period 3 (11:30-17:00, daytime period), and Period 4 (17:00-22:00, night period). On this basis, the load data are transformed into PAA representations which also vary from 0 to 1. Figure No.3 shows the histogram and CDF of PAA representations of the whole information sets. It may be seen that the upper the consumption, the lower the density. For the number of Markov states, we change it from 1 to 6, and then calculate the average recovery error for every case, as shown in The common error drops apace because the variety of states increases. However, it changes little when the number of states is greater than 3. The breakpoints are approximately valued as 0.1 (one-tenth of the maximum consumption) and 0.25 (one-fourth of the maximum consumption) corresponding to 0.333 and 0.667 in CDF respectively. Thus, we divide the amplitude into 3 parts: image a for 0~0.1; image b for zero.1~0.25; and image c for 0.25~1.0. These three states can be defined as absence, passive occupancy, and active occupancy[23]. Then, the electricity consumption knowledge of every individual client will be painted as a symbolic string, like the case in Figure No.3. Then, four Markov model of each customer are modeled for the four periods of the day. We calculated the 2 χ test statistic according to (5) for 6445 customers over four periods. Given the significance level α =0.05, 2 2 2 0.05 ((N 1)) (4 9.48) 8χχα–==. The results show that the electricity consumption of over 99% (6387) of the customers have much larger 2 χ test statistic and show a significant Markov property. C. Clustering for Full Periods to obtain the typical dynamic characteristics of electricity consumption and to segment customers into several groups, CFSFDP is first applied to the full periods. After calculating the dissimilarity matrix following (8), we plot the local density ρ and distance δ of each customer, calculated according to (9) and (10), respectively, in the decision graph, as shown in Figure No.8 We choose the density peak with ρ> 10 and δ> 0.5, where a total of 40 clusters can be obtained, which have been marked with different colors in Figure No.8. To show the distribution of the 6445 customers, we mapped the customers into a 2-D plane according to their dissimilarity matrix by multidimensional scaling (MDS)[30] as shown in Figure No.9. MDS is a very effective dimensional reduction way for visualizing the level of similarity among different objects of a data set. It tries to place each object in N-dimensional space such that the between-object distances area unit preserved as closely as attainable. Every purpose in the plane stands for a customer. Points in the same cluster are marked with the same color. It can be seen that the customers of different clusters are unevenly distributed. Approximately 90% of the customers belong to the 10 larger clusters, whereas the other 10% are distributed in the alternative thirty clusters. During this method, these 6445 customers area unit segmental into totally different teams according to their electricity consumption dynamic characteristics for full periods. Note that the customers in the same cluster has similar electricity consumption behavior dynamics over a certain period instead of similar shape in load profiles. D. Clustering for Each Adjacent Periods Sometimes, we may not be concerned with the dynamic characteristics of full periods and instead concentrate on a certain period of time. For example, to evaluate the demand response potential in high noon peak shaving of every client, the dynamics from amount one to amount two square measure much more important; to measure the potential to follow the change of wind power at midnight, the dynamics from amount four to amount one ought to be emphasized. Thus, it's necessary to conduct customer segmentation for different adjacent periods. illustrates the decision graph and 2-D plane mapping of consumers for the four adjacent periods. It will be seen that the distributions of the customers of the four adjacent periods are shaped like bells, and the proposed clustering technique can effectively address the nonspherically distributed data. Unsurprisingly, the dynamics from Period 2 to Period 3 and from Period 3 to Period 4 show more diversity because people become more active during the day, whereas the dynamics from Period 1 to Period 2 and from Period 4 to Period 1 show less diversity because most people are off duty and go to sleep with less electricity consumption. Taking the dynamics from Period 2 to Period 3 as an example, the six most typical dynamic patterns are shown in Figure No.2. The percent in each matrix stands for the percentage of customers who belong to the cluster. For example, approximately 37% of the customers have very similar electricity consumption dynamics to that of Type_1.

Then, the distortion threshold θ is carefully selected for the adaptive k means method, as a larger threshold leads to poor accuracy, whereas a smaller one leads to little compression. We run 100 cases by varying θ from 0.0025 to 0.25 with steps of 0.0025 and calculate the average compression ratio (CR) of the three distributed sites for each case. The CR is defined as the ratio between the amount of the compressed information and therefore the volume of the first information. Especially, the compressed data refers to local models obtained by adaptive k-means, and the original data refers to the whole objects distributed on each sites: CR No. of local models.

POTENTIAL APPLICATIONS

Different from the traditional load profiling methods which mainly focus on the shape of load profiles, this paper tries to perform clustering on the load consumption change extents and possibilities in adjacent time periods, indicating dynamic features of customer consumption behaviors. The proposed modelling method has many potential applications. For example, on the decision graph obtained by CFSFDP such as Figure No.7 and Figure No.8, we can easily find the objects with small ρi and large iδ , which can be considered as outlier. That is to say, this customer shows great difference on electricity consumption behavior dynamics. However, customers of similar social eco-backgrounds are more likely to have similar electricity consumption behavior dynamics. Thus, we can detect abnormal or suspicious electricity consumption behavior quickly through the decision graph. For another example, future consumption can be simulated through Monte-Carlo from the angle of statistics and probability if the state transition probability matrix is known. Based on the simulated electricity consumption, optimal ToU tariff can be designed. Moreover, entropy based demand response targeting will be further analyzed in this section as an illustration of the applications. It is believed that customers of less variability and heavier consumption are suitable for incentive-based demand response programs like direct load control (DLC) for their predictability for control, whereas customers of greater variability and heavier consumption are suitable for price-based demand response programs, like ToU pricing, for their flexibility to modify their consumption. Note that a N N× state transition probability matrix is essentially a combination of N probability distributions as mentioned before. Obviously, though the dynamic characteristics have been abstracted into 3 3 × matrices as shown in we can make intuitive evaluations on the customers toward demand response targeting by introducing the approach of entropy evaluation to further extract information from the matrices. The variability could be quantified by the Shannon entropy[23] of the state transition matrix Table II shows the entropies of the Markov model in Figure No.9. It can be seen that Type_3 shows the minimum entropy. The 0.994 in the Type_3 matrix means that the Type_3 customers have a greater opportunity to remain unchanged in state c, i.e., the higher consumption level, and are easier to predict. Thus, customers of Type_3 may have a greater potential for an incentive-based demand response during Period 3. However, Type_1 and Type_2 show much higher entropies and have a relative higher consumption level than Type_3, which makes them much more suitable for a price-based demand response. For example, the Type_1 and Type_2 customers have almost the same probability of switching from state c to state b and state c, which is hard to predict, and have more flexibility to adjust their consumption behaviors.

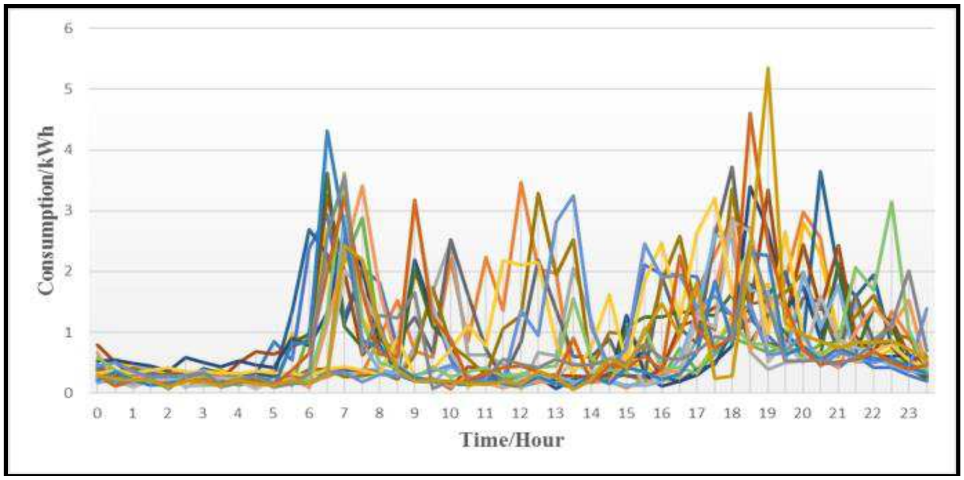

Figure No.1: Analysis diagram

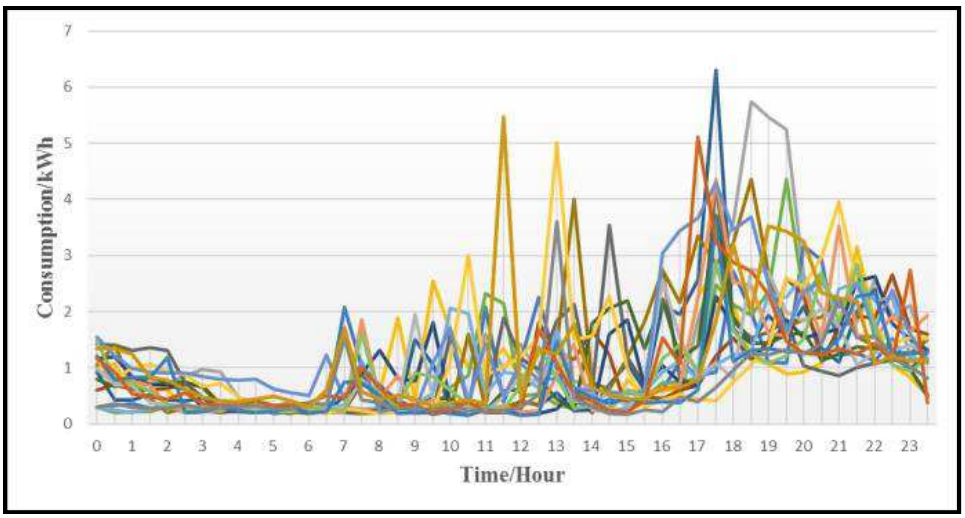

Figure No.2: Analysis per time

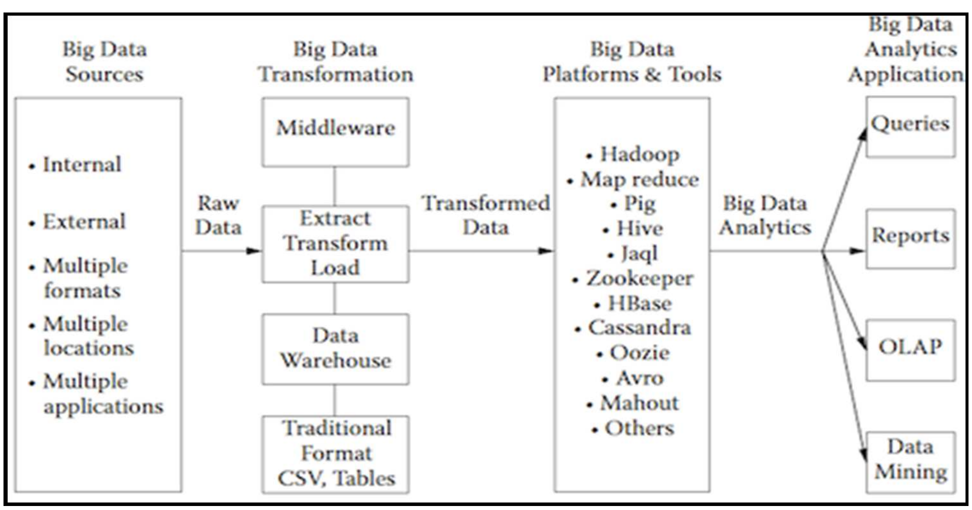

Figure No.3: Data Transformation

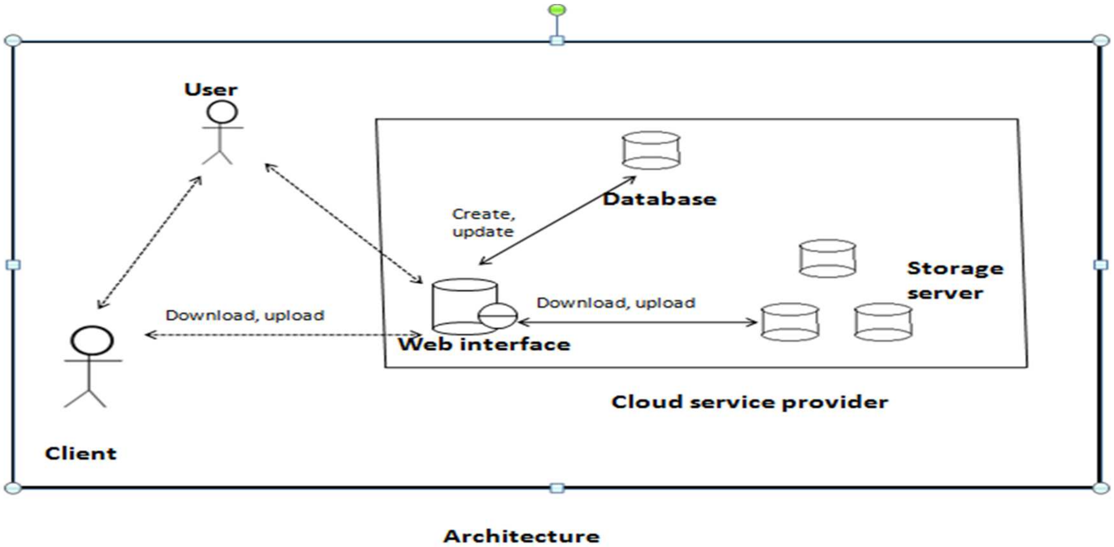

Figure No.4: Usecase diagram

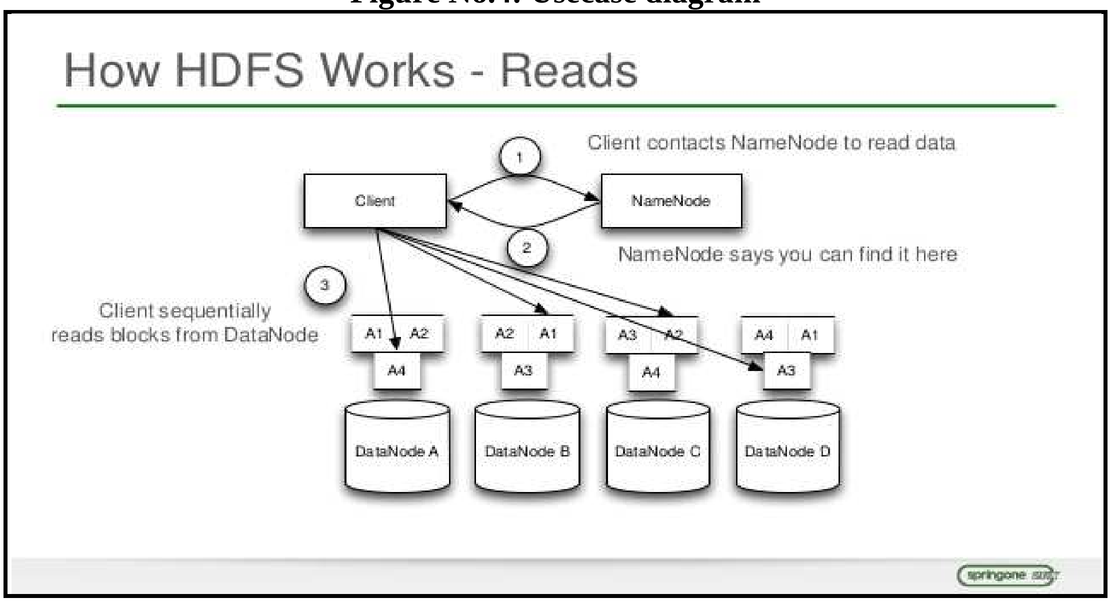

Figure No.5: Workflow diagram

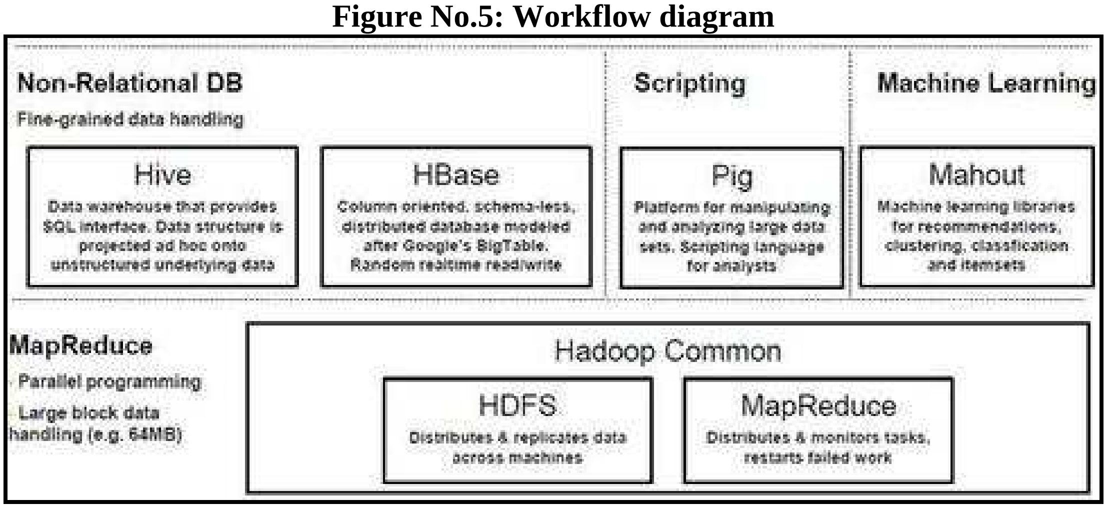

Figure No.6: Database Relationship

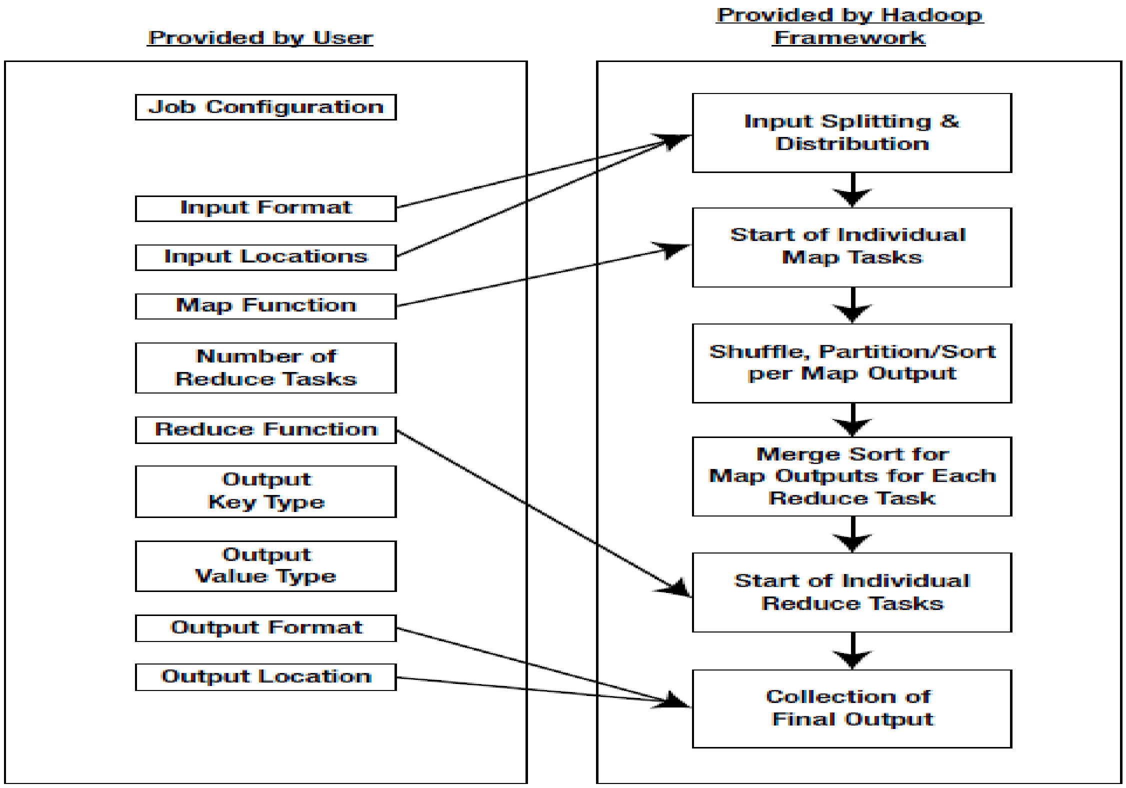

Figure No.7: DB Connectivity

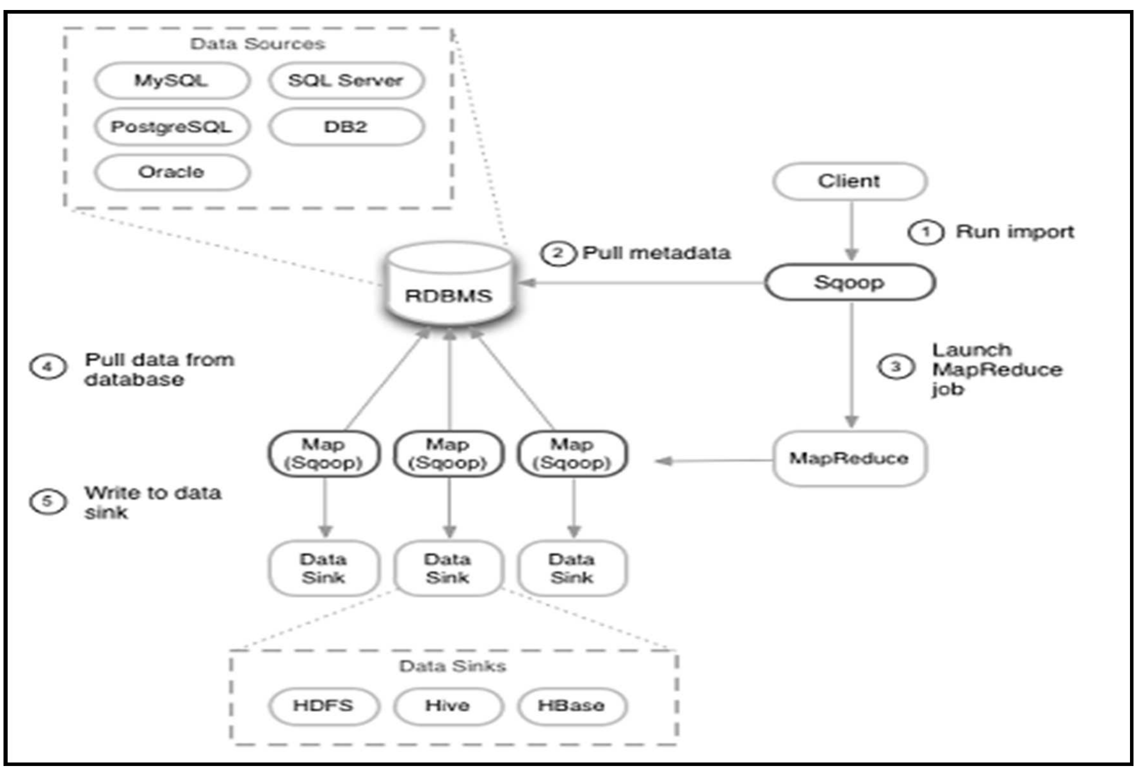

Figure No.8: DB Architecture

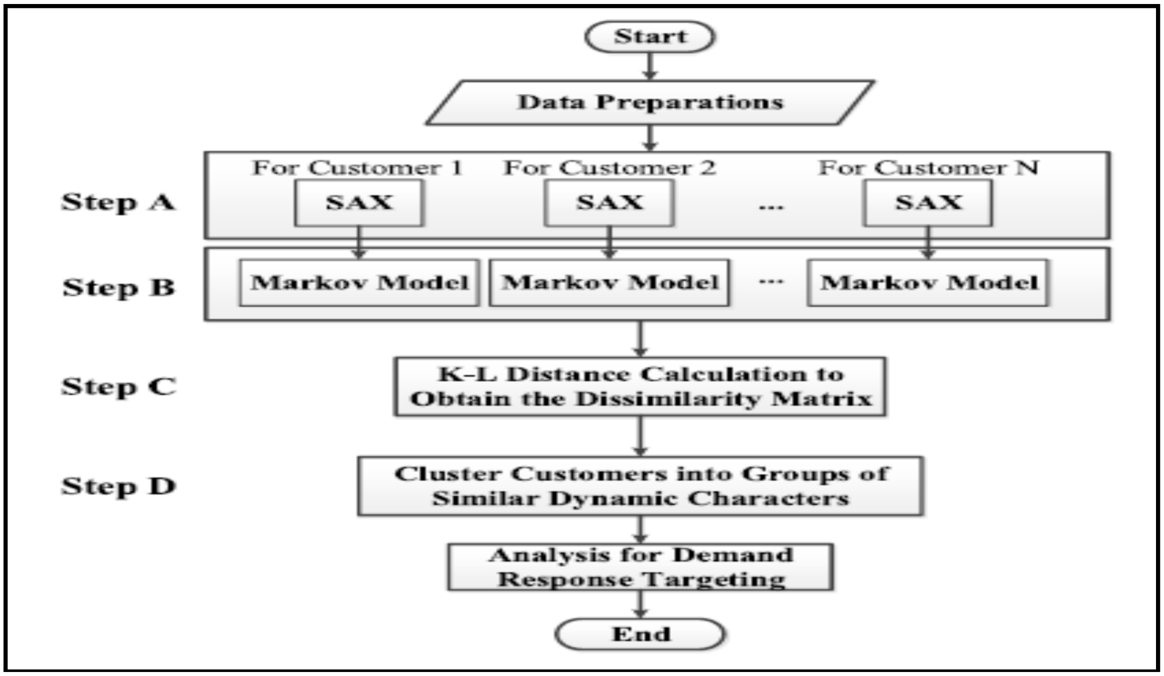

Figure No.9: Flow chart

CONCLUSION AND FUTURE WORK

In this paper, a completely unique approach for the agglomeration of electricity consumption behavior dynamics toward large data sets has been proposed. Different from traditional load profiling from a static prospective, SAX and time-based Markov model square measure used to model the electricity consumption dynamic characteristics of each customer. A density-based clustering technique, CFSFDP, is performed to get the everyday dynamics of electricity consumption and phase customers into different groups. Finally, a time domain analysis and entropy evaluation are conducted on the results of the dynamic agglomeration to spot the demand response potential of each group's customers. The challenges of massive high-dimensional electricity consumption data square measure addressed in 3 ways. First, SAX will scale back and discretize the numerical consumption data to ease the cost of data communication and storage. Second, Markov model are modelled to transform long-run information to many transition matrixes. Third, a distributed agglomeration algorithm is proposed for distributed big data sets. Limited by the data sets, the influence of external factors like temperature, and economy, day type on the electricity consumption isn't thought of in depth in this paper. Future works will focus on feature extraction and data mining techniques combining electricity consumption with external factors.

ACKNOWLEDGEMENT

The authors would like to thank Wenjun Zhu (Guangdong Power Grid Corporation) and Min Luo (Power Research Institute of Guangdong Power Grid Corporation) for providing the insights into the distributed clustering algorithm.

CONFLICT OF INTEREST

We declare that we have no conflict of interest.

BIBLIOGRAPHY

1. USA Department of Energy, Smart Grid/ Department of Energy, http: //energy.gov/oe/ technology-development/smart-grid, 2014.
2. Panapakidis I P, Alexiadis M C and Papagiannis G K. "Load profiling in the deregulated electricity markets: A review of the applications," in European Energy Market

(EEM), *2012 9th International Conference,* 2012, 1-8.

3. Granell R, Axon C J and Wallom D C H. "Impacts of Raw Data Temporal Resolution Using Selected Clustering Methods on Residential Electricity Load Profiles," *IEEE Trans. Power Systems,* 30(6), 2015, 3217-3224.
4. Mahmoudi-Kohan N, Moghaddam M P, Sheikh-El-Eslami M K and Shayesteh E. "A three-stage strategy for optimal price offering by a retailer based on clustering techniques," *International Journal of Electrical Power and Energy Systems,* 32(10), 2010, 1135-1142.
5. Zhang P, Wu X, Wang X and Bi S. "Short-term load forecasting based on big data technologies," *CSEE Journal of Power and Energy Systems,* 1(3), 2015, 59-67.
6. Mahmoudi-Kohan N, Moghaddam M P, x K. Sheikh-El-Eslami M P and Bidaki S M. "Improving WFA k-means technique for demand response programs applications," in Power and Energy Society General Meeting, *PES '09, IEEE,* 2009, 1-5.
7. Leon C, Biscarri F, Monedero I, Guerrero J I, Biscarri J and Millan R. "Variability and Trend-Based Generalized Rule Induction Model to NTL Detection in Power Companies," *IEEE Trans. Power Systems,* 26(4), 2011, 1798-1807.
8. Wang Y, Chen Q, Kang C, Zhang M, Wang K and Zhao Y. "Load profiling and its application to demand response: A review," *Tsinghua Science and Technology,* 20(2), 2015, 117-129.
9. Li R, Gu C, Li F, Shaddick G and Dale M. "Development of Low Voltage Network Templates-Part I: Substation Clustering and Classification," *IEEE Trans. Power Systems,* 30(6), 2015, 3036-3044.
10. Zhou K, Yang Sv and Shen C. "A review of electric load classification in smart grid environment," *Renewable and Sustainable Energy Reviews,* 24(C), 2013, 103-110.
11. Tsekouras G J, Kotoulas P B, Tsirekis C D, Dialynas E N and Hatziargyriou N D. "A pattern recognition methodology for evaluation of load profiles and typical days of large electricity customers," *Electric Power Systems Research,* 78(9), 2008, 1494-1510.
12. Verdu S V, Garcia M O, Senabre C, Marin A G and Franco F J G. "Classification, Filtering, and Identification of Electrical Customer Load Patterns Through the Use of Self-Organizing Maps," *IEEE Trans. Power Systems,* 21(4), 2006, 1672-1682.
13. Chicco G and Ilie I S. "Support Vector Clustering of Electrical Load Pattern Data," *IEEE Trans. Power Systems,* 24(3), 2009, 1619-1628.
14. Piao M, Shon H S, Lee J Y and Ryu K H. "Subspace Projection Method Based Clustering Analysis in Load Profiling," *IEEE Trans. Power Systems,* 29(6), 2014, 2628-2635.
15. Chicco G, Ionel O and Porumb R. "Electrical Load Pattern Grouping Based on Centroid Model with Ant Colony Clustering," *IEEE Trans. Power Systems,* 28(2), 2013, 1706-1715.
16. Chicco G. "Overview and performance assessment of the clustering methods for electrical load pattern grouping," *Energy,* 42(1), 2012, 68-80.
17. Fodor I K. "A Survey of Dimension Reduction Techniques," *Perpinan,* 205(3), 2003, 351-359.
18. Abrahams M and Kattenfeld M. "Two-stage fuzzy clustering approach for load profiling,", in Universities Power Engineering Conference (UPEC), *2009 Proceedings of the 44th International,* 2009, 1-5.
19. Koivisto M, Heine P, Mellin I and Lehtonen M. "Clustering of Connection Points and Load Modeling in Distribution Systems," *IEEE Trans. Power Systems,* 28(2), 2013, 1255-1265.
20. Chicco G, Napoli R and Piglione F. "Comparisons Among Clustering Techniques

for Electricity Customer Classification," *IEEE Trans. Power Systems*, 21(2), 2006, 933-940.

21. Varga E D, Beretka S F, Noce C and Sapienza G. "Robust Real Time Load Profile Encoding and Classification Framework for Efficient Power Systems Operation," *IEEE Trans. Power Systems*, 30(4), 2015, 1897-1904.
22. Zhong S and Tam K. "Hierarchical Classification of Load Profiles Based on Their Characteristic Attributes in Frequency Domain," *IEEE Trans. Power Systems*, 30(5), 2015, 2434-2441.
23. Torriti J. "A review of time use models of residential electricity demand," *Renewable and Sustainable Energy Reviews*, 37(2), 2014, 265-272.
24. Xiao Y, Yang J, Que H. "Application of Wavelet-based clustering approach to load profiling on AMI measurements," in Electricity Distribution (CICED), *2014 China International Conference on, IEEE*, 2014, 1537-1540.
25. Notaristefano A, Chicco G, Piglione F. "Data size reduction with symbolic aggregate approximation for electrical load pattern grouping," *Generation, Transmission and Distribution, IET*, 7(2), 2013, 108-117.
26. Albert A and Rajagopal R. "Smart Meter Driven Segmentation: What Your Consumption Says About You," *IEEE Trans. Power Systems*, 28(4), 2013, 4019-4030.
27. Rodriguez M, Gonzalez I, Zalama E. "Identification of Electrical Devices Applying Big Data and Machine Learning Techniques to Power Consumption Data," *International Technology Robotics Applications Springer International Publishing*, 2014, 37-46.
28. Rodriguez A, ALaio. "Clustering by fast search and find of density peaks," *Science*, 334(6191), 2014, 1492-1496.
29. Lin J, Keogh E, Lonardi and Chiu B. "A Symbolic Representation of Time Series, with Implications for Streaming Algorithms,", In proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, *San Diego, CA*, 2003, 2-11.
30. Kullback S, Leibler R A. "On information and sufficiency," *The annals of mathematical statistics*, 22(1), 1951, 79-86.
31. Jessica Lin, Eamonnkeogh, Stefano lonardi, Bill Chiu. A Symbolic Representation of Time Series, with Implications for Streaming Algorithms, *ACM, New York, USA*, 2003, 2-11.

Please cite this article in press as: Jagadesh S and Mahesh A.A time based Markov model to formulate the dynamics of customers electricity consumption behavior, *International Journal of Engineering and Robot Technology*, 4(1), 2017, 1-12.