International Journal for Modern Trends in Science and Technology
Volume: 03, Special Issue No: 02, March 2017
ISSN: 2455-3778
http://www.ijmtst.com

Analyzing the Behavior of Electricity Consumption Using Hadoop

B.N.K.Uday Kiran[1] | M.Harsha Sainath[2] | M.Shirdi Lal[3] | S.Kishore Babu[4]

[1,2,3,4]Department of IT, Andhra Loyola Institute of Engineering & Technology, Vijayawada, Andhra Pradesh, India.

To Cite this Article
B.N.K.Uday Kiran, M.Harsha Sainath, M.Shirdi Lal and S.Kishore Babu, "Analyzing the Behavior of Electricity Consumption Using Hadoop", *International Journal for Modern Trends in Science and Technology*, Vol. 03, Special Issue 02, 2017, pp. 64-68.

ABSTRACT

In the present day retail market, there are several opportunities for load serving entities which are provided by large volumes of smart data for meters , which improves the knowledge of electricity consumption behaviors of customers by using load profiling instead of focusing on load curves. This paper proposes a unique approach for clustering the electricity consumption behavior dynamics such as transitions and relations between them in eventual periods. First, to downsize the scale of data a symbolic aggregate approximation (SAX) is performed for each distinct customer and to model the electricity consumption dynamic, transforming the large data set of load curves to several state transition matrixes by using a time-based Markov model is applied. Second, to obtain the dynamics of consumption behavior a clustering technique by Fast Search and Find of Density Peaks (CFSFDP) is mainly carried out, with the distinction between any two consumption patterns measured by the Kullback–Liebler (K-L) distance, and to classify the customers into several clusters. To tackle the challenges of big data, 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. with the refinement between any two utilization designs measured by the Kullback–Liebler (K-L) remove, and to arrange the clients into a few bunches. To handle the difficulties of enormous information, the CFSFDP method is coordinated into a gap and-overcome approach toward huge information applications. A numerical case checks the adequacy of the proposed models and methodologies.

I. Introduction

Nations around the globe have set forceful objectives for the rebuilding of monopolistic power framework towards changed markets particularly on the request side. In a focused retail advertise, stack serving elements (LSEs) will be created in incredible numbers. Having a superior comprehension of power utilization designs and acknowledging customized control administrations are viable approaches to upgrade the aggressiveness of LSEs. In the interim, savvy frameworks have been upsetting the electrical era and utilization through a two-route stream of force and data. As a vital data source from the request side, progressed metering framework (AMI), has increased expanding prominence around the world; AMI permits LSEs to get power utilization information at high recurrence, e.g., minutes to hours. Huge volumes of power utilization information uncover data of clients that can possibly be utilized by LSEs to deal with their era and demand assets productively and give customized benefit. Stack profiling, which alludes to power utilization practices of clients over a particular period, e.g., one day, can help LSEs see how power is really utilized for various clients and get the clients' heap profiles or load designs. Stack profiling assumes a fundamental part in the Time of Use (ToU) duty outline, nodal or client scale stack estimating, request reaction and vitality proficiency focusing on, and non-specialized misfortune (NTL) discovery.

The center of load profiling is grouping which can be characterized into two classifications: coordinate bunching and circuitous bunching. Coordinate grouping implies that bunching techniques are connected straightforwardly to load information. Leading up to now, there are an extensive number of grouping systems that are broadly considered, including k-implies , fluffy k-implies , various leveled bunching , self-sorting out maps (SOM) , bolster vector grouping, subspace grouping , insect settlement grouping and so on. The execution of each grouping method could be assessed and evaluated utilizing different criteria, including the bunching scattering marker (CDI), the disseminate list (SI), the Davies-Bouldin list (DBI), and the mean file sufficiency (MIA). .

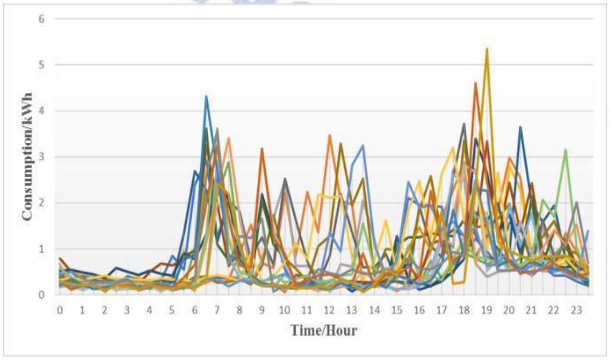
The storm of power utilization information with the broad and high-recurrence accumulation of brilliant meters presents awesome difficulties for information stockpiling, correspondence and investigation. In this unique situation, measurement decrease techniques can be successfully connected to diminish the extent of the heap information before grouping, which is characterized as aberrant bunching. Such bunching can be classified into two sub classifications, highlight extraction-based grouping and time arrangement based bunching. Highlight extraction which changes the information in the high-dimensional space into a space of less measurements, is frequently used to diminish the size of the info information. Key segment investigation (PCA) is a much of the time utilized direct measurement diminishment strategy. It tries to hold the greater part of the covariance of the information highlights with the least manufactured factors. Some nonlinear measurement diminishment strategies including Sammon maps, curvilinear segment examination (CCA) , and profound learning have additionally been connected to power utilization information. Additionally, as power utilization information is basically a period arrangement. An assortment of develop scientific strategies, for example, discrete Fourier change (DFT) , discrete wavelet change (DWT) , typical total guess (SAX) , and the concealed Markov demonstrate (HMM) have been talked about in the writing. These techniques are fit for lessening the dimensionality of time arrangement and of keeping up a portion of the first character of the electrical utilization information.

The current reviews on load profiling for the most part concentrate on individual vast modern/business client, medium or low voltage feeder, or a mix of little clients, stack profiles of which shows substantially more consistency . It ought to be noticed that in spite of the fact that these dynamic qualities are constantly "deluged" in a blend of clients, they could be depicted by a few common load designs. Be that as it may, with respect to private clients, no less than two new difficulties will be confronted. One test is the high assortment and fluctuation of the heap designs. As demonstrated by Fig 1, there are clear contrasts in the power utilization examples of the two inhabitants. Crest loads have distinctive amplitudes and happen at various circumstances of day, for instance. Power utilization designs additionally fluctuate once a day notwithstanding for a similar client. For this situation, a few normal day by day stack examples are not sufficiently fine to uncover the genuine utilization practices. The day by day profile ought to be disintegrated into all the more fine-grained sections, which are powerfully changed and recognized. In addition, as the utilization conduct of a particular client is basically a state-needy, stochastic process, it is imperative to investigate the dynamic attributes, e.g., exchanging and keeping up, of the utilization states and the relating probabilities. The other test is that of "huge information". Considering the high recurrence and dimensionality of the information contained in the heap bends, informational collections in the multi-petabyte range will be investigated. Conventional grouping systems are dubious to be executed in a "major information world".

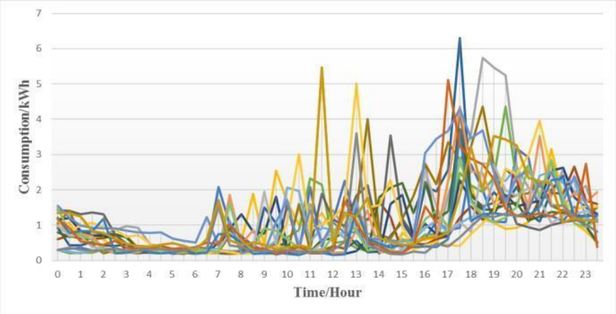
To handle these two difficulties, this paper actualizes a period based Markov model to detail the progression of clients' power utilization practices, considering the state-subordinate attributes, which shows that future utilization practices would be identified with the present states. This supposition is sensible as different power utilization practices would keep going for various timeframes before being fit for change, as could be disconnected from verifiable exhibitions. The moves and relations between utilization practices, or rather utilization levels, in contiguous periods are alluded to as "flow" in this paper. These progression have been demonstrated by Markov show in a few works . In any case, few papers consider the elements as a variable for grouping. Profiling of the progression could give valuable data to understanding the utilization examples of clients, gauging the utilization drifts in brief eras, and recognizing the potential request reaction targets. Additionally, this approach defines the substantial informational index of load bends as a few state move frameworks, incredibly lessening the dimensionality and scale.

Notwithstanding the Markov display, this paper tries to address the "information storm" issue in three different ways. To start with, applying SAX to change the heap bends into a typical string to decrease the storage room and facilitate the correspondence movement between brilliant meters and server farms. Second, an as of late revealed successful bunching procedure by Fast Search and Find of Density Peaks (CFSFDP) is initially used to profile the power utilization practices, which has the upsides of low time multifaceted nature and heartiness to commotion focuses. The elements of power utilizations are depicted by the differences between each two utilization designs, as measured by the Kullback–Liebler (K-L) remove. Third, to handle the difficulties of enormous and scattered information, the CFSFDP strategy is incorporated into a partition and-overcome way to deal with further enhance the productivity of information preparing, where versatile k-means is connected to get the delegate clients at the neighborhood destinations and an altered CFSFDP technique is performed at the worldwide locales.

The approach could be further connected toward enormous information applications.



(a)Fig.1



(b)Fig.2

At long last, the potential utilizations of the proposed strategy to request reaction focusing on, irregular utilization conduct recognizing and stack determining are investigated and examined. Particularly, entropy investigation is directed in light of the grouping results to assess the inconstancy of utilization conduct for each bunch, which can be utilized to measure the capability of value based and motivating force based request reaction.

The commitments of this paper are as per the following:

1） Time-based Markov model is connected to define the power utilization conduct elements rather than the state of day by day stack profiles.

2） Customer division is performed by a high-productive grouping calculation named CFSFDP which is vigorous to commotion and need no emphases.

3） A conveyed bunching structure joining versatile k-means and CFSFDP is proposed to handle the expansive and dispersed informational collection.

4） The use of the proposed displaying strategy and profiling calculation are examined and talked about.

Whatever remains of the paper is sorted out as takes after: In Section II the essential approach of grouping of power utilization conduct flow is presented. In Section III, a separation and-overcome circulated bunching calculation for huge informational indexes is proposed. In Section IV contextual investigations and some examination for request reaction focusing on and dispersed grouping are led in view of open information from Ireland.

The proposed philosophy for the dynamic disclosure of the power utilization can be partitioned into six phases, as appeared in Fig.2. The principal organize directs some heap information arrangements, including information cleaning and load bend standardization. The second stage decreases the dimensionality of the heap profiles utilizing SAX. The third stage defines the power utilization elements of every individual client using time-based Markov display. The K-L separation is connected to gauge the contrast between any two Markov model to get the separation grid in the fourth stage. The fifth stage plays out an adjusted CFSFDP grouping calculation to find the normal progression of power utilization. At long last, the consequences of the examination of the request reaction focusing on are gotten in the 6th stage. The subtle elements of the initial five phases will be presented in the accompanying, and the request reaction focusing on examination part will be further clarified for the situation considers

II. BASIC METHODOLOGY

The proposed philosophy for the dynamic revelation of the power utilization can be isolated into six phases, as appeared in Fig. 2.

The principal organize leads some heap information arrangements, including information cleaning and load bend standardization. The second stage diminishes the dimensionality of the heap profiles utilizing SAX. The third stage defines the power utilization progression of every individual client using time-based Markov demonstrate. The K-L separation is connected to gauge the distinction between any two Markov model to acquire the separation framework in the fourth stage. The fifth stage plays out a changed CFSFDP bunching calculation to find the common elements of power utilization. At last, the consequences of the investigation of the request reaction focusing on are acquired in the 6th stage. The points of interest of the initial five phases will be presented in the accompanying, and the request reaction focusing on examination part will be further clarified for the situation contemplates.

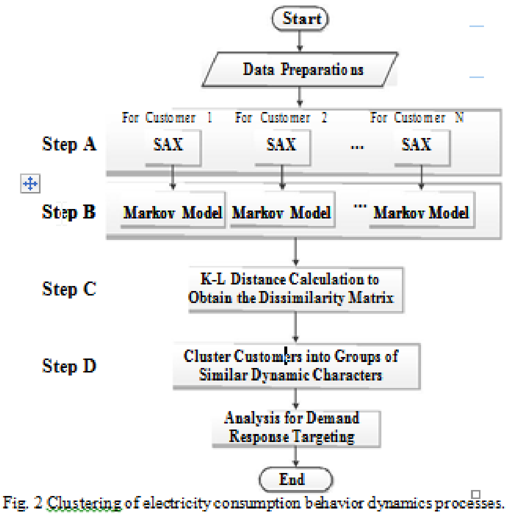

Start
Data Preparatio ns
For Customer 1
For Customer 2
For Customer N
Step A
SAX
SAX
...
SAX
Step B
Markov Model
Markov Model
...
Markov Model
Step C
K-L Distance Calculation to Obtain the Dissimilarity Matrix
Step D
Cluster Customers into Groups of Similar Dynamic Characters
Analysis for Demand Response Targeting
End

Fig. 2 Clustering of electricity consumption behavior dynamics processes.

A. Data Normalization

Information arrangements including information cleaning is not the subject of this paper and won't be talked about. To make the heap profiles equivalent, the standardization procedure changes the utilization information of subjective value x1 , x2 ,to the scope of (0,1).

This technique is decided for no less than three reasons. To start with, it can debilitate the effect of anomalous days with basic pinnacles infusions. Second, it can give stack shapes little impact from day by day or regular changes in the greatest qualities. Third, it can sift through the base load, which has little impact on request reaction and hold, for the fluctuant part, which demonstrates more noteworthy potential sought after reaction.

B. SAX for Load Curves

Where, xi and xi indicate the real and standardized power utilization at time i; x and x mean the base and most extreme utilization over H periods respectively. It ought to be noticed that the standardization is performed once a day rather than over whole periods. This technique is decided for no less than three reasons. To start with, it can debilitate the effect of anomalous days with basic pinnacles infusions. Second, it can give stack shapes little impact from day by day or regular changes in the greatest qualities. Third, it can sift through the base load, which has little impact on request reaction and hold, for the fluctuant part, which demonstrates more noteworthy potential sought after reaction.

C. Time-based Markov Model

In the event that we need to anticipate the pattern or level of power utilization for every client, we may make full utilization of their over a wide span of time states. In the event that the future utilization level or state depends just on the present state, it is known as a Markov property and can be displayed by a Markov chain. Different Markov models have been connected to load guaging.

For a typical string with N images, discrete Markov show with N comparing states can be connected to demonstrate the dynamic qualities of their utilization levels. In any case, clients have diverse element qualities at various periods for their normal schedules each day. In this manner, time-based Markov model is connected to define the qualities.

we can be reasonably confident that the

electricity consumption of customers has a Markov property.

D. Distance Calculation

Uniqueness and separate estimation is a crucial issue in bunching. There exist numerous approaches to process the separations between two grids, for example, 1-standard separation and 2-standard separation (Euclidean separation). Be that as it may, not quite the same as general networks, a N state move likelihood framework basically comprises of N likelihood conveyances, where each column (e.g., the ith push) compares to a probabilistic dissemination of the condition of the following time frame at the present state (e.g., the ith state). K-L separation is a powerful approach to measure the disparity between two probabilistic circulations.

E. CFSFDP Algorithm

CFSFDP is an as of late announced bunching calculation that can viably perceive groups paying little mind to their shape with a sensible suspicion that the cluster centers must have a higher nearby thickness and moderately bigger separation to the focuses of higher thickness.

For an informational index, the neighbors can be perceived by a delicate edge like the Gaussian portion work or a hard edge as characterized in above area. To lessen the calculation multifaceted nature for huge informational collections, we utilize the hard edge to compute the neighborhood thickness.

III.CONCLUSION

In this paper, a novel approach for the clustering 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 are utilized to model the electricity consumption dynamic characteristics of each customer. A density-based clustering technique, CFSFDP, is performed to discover the typical dynamics of electricity consumption and segment customers into different groups. Finally, time domain analysis and entropy evaluations are conducted on the result of the dynamic clustering to identify the demand response potential of each group's customers. The challenges of massive high-dimensional electricity consumption data are addressed in three ways. First, SAX can reduce and discretize the numerical consumption data to ease the cost of data communication and storage. Second, Markov model are modelled to transform long-term data to several transition matrixes. Third, a distributed clustering algorithm is proposed for distributed big data sets. Limited by the data sets, the influence of external factors like temperature, day type, and economy on the electricity consumption is not considered in depth in this paper. Future works will focus on feature extraction and data mining techniques combining electricity consumption with external factors.

References

[1] USA Department of Energy, Smart Grid / Department of Energy, http://energy.gov/oe/technology-development/smart-grid, 2014.

[2] I. P. Panapakidis, M. C. Alexiadis and G. K. Papagiannis, "Load profiling in the deregulated electricity markets: A review of the applications," in European Energy Market (EEM), 2012 9th International Conference on the, 2012, pp. 1-8.

[3] R. Granell, C. J. Axon and D. C. H. Wallom, "Impacts of Raw Data Temporal Resolution Using Selected Clustering Methods on Residential Electricity Load Profiles," IEEE Trans. Power Systems, vol. 30, pp. 3217-3224, 2015.

[4] N. Mahmoudi-Kohan, M. P. Moghaddam, M. K. Sheikh-El-Eslami, and E. Shayesteh, "A three-stage strategy for optimal price offering by a retailer based on clustering techniques," International Journal of Electrical Power & Energy Systems, vol. 32, pp. 11351142, 2010.