JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

TO FIND THE IMPACT OF CLUSTERED DATA COMPARED WITH AGGREGATED DATA FOR PRICING SCHEME FOR DEMAND RESPONSE

Sunil Kumar, Er Gurpreet Kaur
M.Tech Student, Guru Kashi University, Talwandi Sabo
Assistant Professor, Guru Kashi University, Talwandi Sabo
sunilvashist08@gmail.com
ergurpreet88@ gmail.com

Abstract- The exponentially rising technology and installation of smart meters open doors to new opportunities and challenges for data analysis in power systems. In the power system market, the rate of tariff is decided considering a lot of factors. These factors vary for the industrial, commercial, and residential sectors. In this thesis, the main focus is given to the residential sector. In the residential sector, the main deciding factor for setting the tariff is the load profile. In this paper, the real-time dynamic price is being selected. The main challenge is of aggregation of the various numbers of customers and choosing a single profile to decide the dynamic price. Earlier the mean of the residential customer was been considered to decide the tariff plan, which results in poor demand-side management and loss of information. But with the technology available today provides the opportunity to further improve pricing by customer-specific load profiles. In this thesis, the difference between load profile extracted by averaging the data and profiles extracted using k means clustering is been shown for a single customer.

Keywords:

INTRODUCTION

Clustered data appear when the values from the study can be categorized into a frequency of unique groups, mention as clusters. Every cluster contains many observations, giving the figures a "nested" structure, with single observations embedded within the cluster.

Clustered figures are extremely plentiful. All the occasions we experience and that we might wish to find, recognize, and affect upon have limited and quantifiable time periods. It, therefore, obey that the single figures points related to every instance of such incident are clustered W.R.T time. Several incidents related wit lustered figures, such as medical and upheaval, can be highly significant, and it is, therefore, necessary to recognize them as precisely as possible. The identifications described the use of CMAs greatly upgrade the ease and precision with which one can

recognize clusters of verification for events of interest. The "smoothing" of the figure that the moving mean procedure does within limits more than merely upgrade the aspect of analog depictions of the figure. The moving means also increment the accuracy with which one can recognize and characterize the affair that the data were collected to reveal.

Although a diversity of statistical measures and specimen sizes can be used to calculate moving means, the present CMA produced the best results.

Aggregate data mention numerical or non-numerical details that are

(1) Collected from many sources measures, variables, or individuals.

(2) Assemble into figure summaries, commonly for the purposes of announcing or analytical survey for example examining drift, making contrast, or divulging statistics and intuitions that would not be obtained when figure elements are shown in isolation. For example, figure about whether single students graduated from college can be lumped that is, collate and abridge into an individual graduation rate for a graduating college, and annual graduation mean can then be clustered into graduation mean for districts.

While most piled education facts are numerical e.g., graduation and dropout value, mean standardized-test scores for a collage of district, the mean amount of finance spent /student in a state, etc. it's both possible and regular to cluster non-numeric figures. For example, masters, tutee, and parents in a college district maybe look over a topic, and the figures and comments from those observed could then be "collected" into a report that observed singles generally think and feel about the matter. Figures gathered during polls, interviews, and focus groups can be clustered.The method of aggregate figures and how it will be used in power system, consider a grid with an enrollment of 500 households, which means the grid maintains 500 households records, every of which contains a wide type of facts about the enrolled households for example, first and last name, home address, date of birth, gender identification, race, date and period of enrollment, the type taken and completed, etc. Once or twice a year, the household's district may be required to submit household reports to their state department of the grid. Every grid in the district will then compile a report that documents the number of households currently enrolled in the grid and in every grade level, which requires administrators to maintain data from all their individual households' records to produce the enrollment reports. The district now has clustered enrollment facts about the households attending its grids. Over the next five years, the household's district could use these annual reports to analyze increment or decreases in district-wide enrollment, enrollment at every household, or enrollment at every grade level. The district could not, however, analyze whether there have been increments or decreases in the enrollment of grids based on the aggregate facts it obtained from its households. To produce a report showing distinct enrollment trends for unique races. The district domestic would then need to disconnect the enrollment facts by racial subgroups.

1.1 Clustered v/s disaggregated facts

To aggregate values is to relate and summarize values; to disaggregate values is to break down aggregated values into component parts or smaller units of values. While this distinction between cluster and disaggregated values may appear straightforward, there is a shade worth talking about here: a lot of "disaggregated" values in education are literal values that have been technically clustered, at some level, from records preserved on single domestic. For example, values rates are hugely considered to be "clustered values," while values rates reported for unique subgroups of domestic—say, for domestic of unique rates and nationalities are typically reviewed to be "disaggregated data." Yet to make reports that disaggregate values rates by race and ethnicity, data on single domestic actually has to be "clustered" to make summary value rates for unique racial subgroups. Most likely, this explanation between clustered and disaggregated values get up the cause, historically, only clustered values on district-wide, or statewide power system performance was ready. When looking over or reporting on topics like clustered value or disaggregated value, it is essential to find precisely how the terms are being used in a special context.

II.PROBEM DEFINING

To perform demand side management, it is the most crucial step to perform the demand response is to set the pricing scheme. The pricing scheme of the residential customer is very much influenced by the customer's load profile. So, previously the aggregation is done by averaging the load profiles of all the customers. But, today a lot of technology is available to extract more consumer-specific load profiling. Therefore, the impact of the clustering should be analyzed. Therefore, based on this the following objectives are formulated:

The price at which the electrical power is been generated by the generating companies is different at different times. It highly depends on the demand at that time the price at which the energy is been sold at the spot electricity market. Therefore, to increase the participation of the customers in demand response, it is necessary that a better pricing scheme should be chosen, such that the customer is tempted to change the consumption pattern according to the need of the utility. Therefore, the first object is to study and analyze the different types of pricing schemes that are available for demand response.

2.2 Objectives

Based on the above discussion the objectives of this paper are:

1. To study and analyze different types of pricing schemes for demand response.
2. To study and analyze different techniques for clustering customers' load profiles based on their patterns.
3. To study and analyze the difference between cluster class representatives and normally aggregated data.
4. To study and analyze the impact of the pricing scheme on clustered and aggregated data.

III Research Methodology

In this thesis a particular customer has been selected from the dataset of 200 households. The customer with maximum consumption has been selected as it can impact the pricing scheme the most. The pricing scheme of the real-time tariff is generally considered as parallel to the load profile, as shown in fig 1.

As in real-time pricing, the aim is to flatten the load profile, so when the load consumption is more the price is kept higher to decrease the power consumption, and where there is low consumption the prices should also be below so as to increase the load consumption. So, these pricing techniques tempt the consumers to change the load profile to a flat curve

Fig1: Real time load and pricing of New York State

V. Result &Discussion

The dataset of 200 households is chosen and the dataset is converted from 10 to one hour. Then maximum load consuming households is selected which is household number 92 with the mean consumption of 26786.58 Watts Hour. The load consumption profile for household 92 is shown in Table 2.

Feature	Value
MEAN	26786.581586
STD	7700.659971
MIN	12515.500000
25%	21278.250000
50%	25234.800000
75%	30773.125000
MAX	67131.300000

The table shows the variability of the consumption pattern of the household 92. The mean consumption is shown below in figure 2.

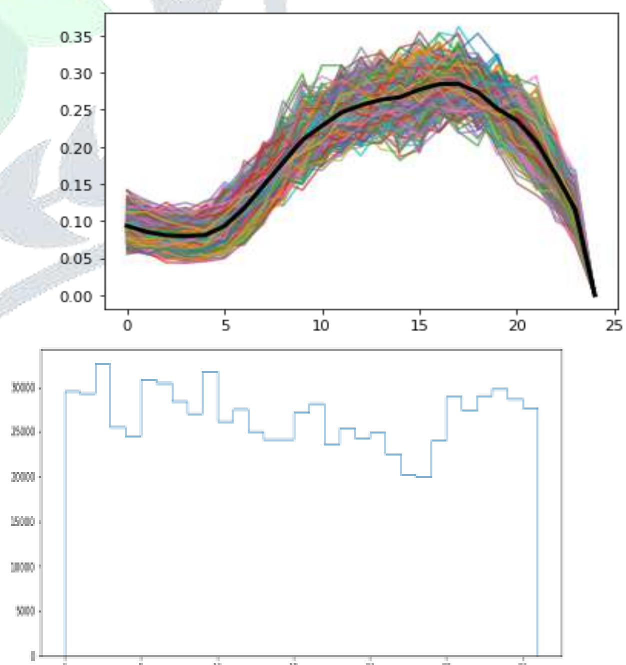


Figure 2: The Mean Consumption of The Household 92 For One Month

Then the clustering is performed on the dataset of a month for household number 92. The elbow method is chosen for selecting the number of clusters. Figure 3 shows the graph from elbow method which is plotted between the sum of square error and number of clusters.

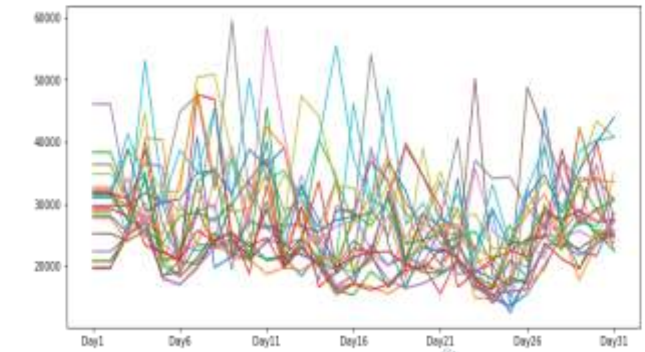


Fig 3: Jan Month Consumption of One Month for Customer 92

If we calculate the Euclidean distance between the self of 92nd house's 9 clusters and the means on the scale of 0 to 1. In this, the distance value total is 0.654 and the mean is 0.13. The value is more than 0.5 therefore it is a considerable difference.

Therefore, if we cluster all the 200 households then we will get 5 clusters according to the elbow method.

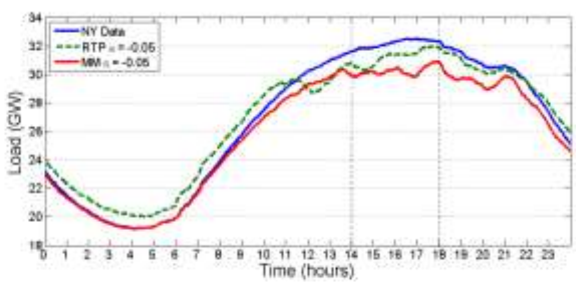


Fig 5. Four Cluster Of 200 Households with Respect to Normalized Load Profile Shapes

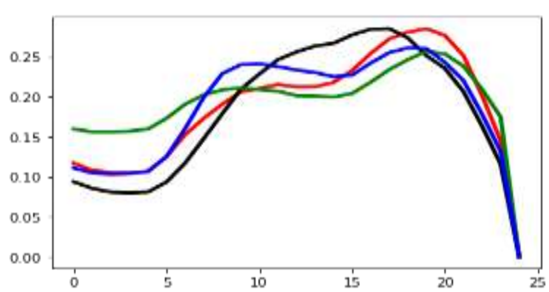


Fig 6. Four Class Represent Of 200 Households with Respect to Normalized Load Profile Shapes

Days	
0	.12
5	.11
10	.21
15	.24
20	.27

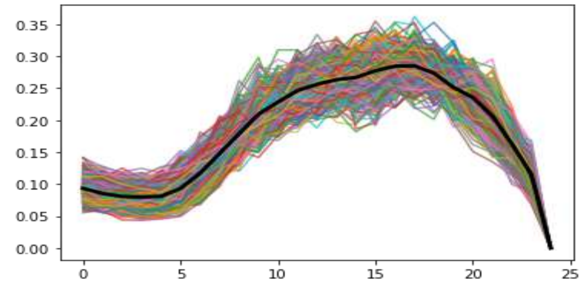
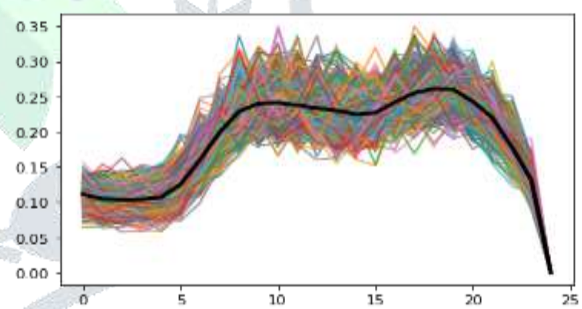
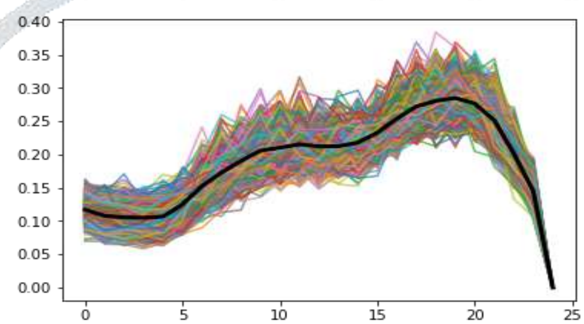


Table 3. Mean Of 200 Households

As in the figure above the difference bet, 9 clusters, and the mean load profile has been calculated. As there is a clear difference between clusters and the mean profile which was considered as the load profile to construct the pricing scheme for the utility. Therefore, it was resulting in less demand response efficiency. Therefore, if the pricing scheme is calculated using each cluster. Each cluster will have a different pricing scheme and hence an improved demand response. The Euclidean distance is even more than 0.354 on a 0-1 scale which is considerable. Therefore, it was better to construct the pricing scheme on the clustered platformer.





VI.Conclusion & Future Scope

The maximum consumption household is been selected for identifying the difference between the aggregation of data by taking the average and identifying clusters. The K- means algorithm has been used for identifying the clusters. The Elbow method is used to find the optimal number of clusters. However, no visible elbow has been formed so the number of the cluster that has been chosen is nine. The above shown is the nine different clusters and the days included in them. However, it can be seen that these clusters are significantly different from each other. It can be judged that taking an average of this data will definitely provide the wrong information and loss of data.

In the future scope, the nature of consumption of the consumer should be further studied and observation should be made so that better prediction and clustering can be formed. Better algorithms including the customer's nature should be evolved.

REFERENCES

[1] J.Luo (VSTLF), "Real-time anomaly detection for very short-term load forecasting," SGEPRI, 2018.

[2] M. Yue, "Anomaly Detection Based on Long Short-Term Memory Neural Network," IEEE, 2017.

[3] A. lazaris, "An LSTM Framework For Modeling Network Traffic," IEEE, 2019.

[4] v. k. prasanna, "A short-term building cooling load prediction method," 2019.

[5] P. Zhao, "Advanced correlation-based anomaly detection method for predictive maintenance," IEEE, 2017.