PREDICTING ELECTRICITY LOAD PATTERN OF CUSTOMERS USING CLUSTERING TECHNIQUE

[1]RUPALI. A. MESHRAM, [2] PROF. A. V. DEORANKAR, [3] DR. P. N. CHATUR

[1]M.Tech. Student, Department Of Computer Science and Engineering, Government College Of Engineering, Amravati, Maharashtra
[2] Asst. Professor, Department Of Computer Science and Engineering, Government College Of Engineering, Amravati, Maharashtra
[3] Head Of The Department Of Computer Science and Engineering, Government College Of Engineering, Amravati, Maharashtra

meshram.rupali@gcoea.ac.in, deorankar.anil@gcoea.ac.in, chatur.prashant@gcoea.ac.in

ABSTRACT: *This paper shows the prediction of electricity demand for future years is an essential step in resource planning and managing tariff of consumers. For clustering the huge amount of data k-means clustering and extended k-means clustering algorithms are used. Precise load forecasting helps the electric utility to make unit commitment decisions, reduce spinning reserve capacity and schedule device maintenance plan properly. Besides playing a key role in reducing the generation cost, it is also essential to the reliability of power systems. The results demonstrate that the proposed method is efficient for assigning Load Profile to the consumers and also shows that the energy consumption can be clustered not only based on the load pattern but also load value. This paper shows work on extended k-means clustering algorithm which helpful for clustering same type of customers into one cluster.*

Keywords— Classification, Load Pattern Analysis, Load Curve, Clustering, Typical Load Profile.

I: INTRODUCTION

An essential step of resource planning in electricity markets is assuring that there will be sufficient resources to meet future demand. Building capacity is costly and takes time. However, the economic consequences of an electricity shortage may be severe. Thus, when setting capacity targets the regulator must balance the cost of excess capacity against the cost of shortage. The question we address here is how can we get an accurate prediction of the future demand. A common approach in many markets is to survey the distribution companies about their predictions of future demand. Although some markets try to provide incentives for the distribution companies to make accurate estimates, such as in Brazil, many other markets provide ambiguous incentives. As a result, distribution companies may be motivated to over or under estimate the future demand. Most of the world's electricity delivery system or "grid" was built when energy was relatively inexpensive.

The result is an inefficient and environmentally wasteful system that is a major emitter of greenhouse gases, consumer of fossil fuels, and not well suited to distributed, renewable solar and wind energy sources. In addition, the grid may not have sufficient capacity to meet future demand.

Several trends have combined to increase awareness of these problems, including greater recognition of climate change, commitments to reduce carbon emissions, rising fuel costs, and technology innovation.

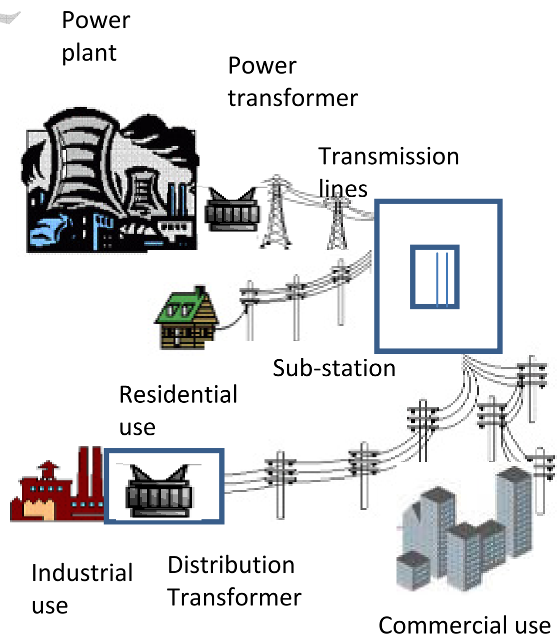

Fig.1Generation and distribution of electricity

In the former case, competing generators offer their electricity output to retailers and, in the latter, end-use customers choose their supplier from competing electricity retailers. Please note that either markets differ from their more traditional counterparts because energy cannot be stored. Consequently, all players are forced to work with consumption prognoses, which, as one may think, creates a number of risks. In this scenario, last-mile electricity customers now have the possibility of choosing their retailer: selecting the most convenient one or, directly going for the worst, will definitively make a difference on the energy bill. Moreover, finding a suitable retailer is not an easy task due to many reasons and this aspect has drawn quite an static electricity market.

Electricity is generated at various kinds of power plant by utilities and independent power producers. Figure 1 shows the generation and distribution of electricity. At prime mover, typically the force of water, steam, or hot gasses on a turbine, spins an electromagnet, generating large amounts of electrical current at a generating station. At transmission side the current is sent at very high voltage (hundreds of thousands of volts) from the generating station to substations closer to the customers. In primary distribution, electricity is sent at mid-level voltage (tens of thousands of volts) from substations to local transformers, over cables called feeders, usually 10-20 km long, and with a few tens of transformers per feeder. Feeders are composed of many feeder sections connected by joints and splices. In secondary distribution phase electricity sends at normal household voltages from local transformers to individual customers.

Load forecasting is necessary for economic generation of power. Load serving entities use load forecasts for system security, to schedule generator maintenance, to make long-term investments in generation, and to plan the most cost-effective merit order dispatch. Over the last decade, as electricity markets have deregulated, the importance of load forecast accuracy has become even more evident. Without an optimal load forecast, utilities are subject to the risk of over- or under- purchasing in the day-ahead market. While an entity can buy or sell power in the real time market to correct for forecast inaccuracy, it comes at the expense of higher real time prices. The aim is to classify the load pattern of different types of customers. Conducting load pattern analysis is an important task in obtaining typical load profiles (TLPs) of customers and grouping them into classes according to their load characteristics. Even if the customer information needed in the classification is correct, some of the customers can simply have such an irregular behaviour pattern that they do not fit in any of the predefined customer class load profiles. The predefined customer class load profiles also include some inaccuracy due to geographical generalization. The most widespread customer class load profiles are created to model the average Finnish electricity consumption. They do not take into account the regional differences in electricity consumption, which originate from different climate conditions and socio-economic factors. The objective is utilization of electricity, developing Tariff on different types of electricity customers, Selection of generators. In the approach, all load curves of customers are first clustered with the clustering algorithms under a serial given number of clusters. When we are using clustering techniques to obtain the load patterns of electricity customers, choosing a suitable clustering algorithm and determining an appropriate cluster number are always important and difficult issues.

Many methods or techniques for clustering load curves have been proposed in the literature. Some clustering methods are: k-means [8], [9], modified follow-the-leader [3], [4], average and Ward hierarchical methods [5], fuzzy c-means (FCM) [9], statistic-fuzzy technique [10], the self-organizing map (SOM) [9], [11], support vector machines (SVM) [2], and extreme learning machine [1]. Some hybrid techniques [12] have been proposed to improve the clustering effect.

The purpose of this paper is to generate typical load profile and classify electric load using k-means clustering and modified k-means clustering. Researchers had worked on to decide which clustering algorithm is best for classification of load pattern analysis for different types of electricity customers. They compared most of the clustering algorithms.

2. ARCHITECTURE OF FORECASTING OF ELECTRICITY

In this work, we focus on the k-means clustering algorithm and extended k-means clustering algorithm. Figure 2 shows the system diagram of generating load profile of electricity customer.

Data cleaning routines work to “clean” the data by filling in missing values, smoothing noisy data, identifying or removing outliers, and resolving inconsistencies. If users believe the data are dirty, they are unlikely to trust the results of any data mining that has been applied to it. Also, dirty data can cause confusion for the mining procedure, resulting in unreliable output. But, they are not always robust. Following two method shows easy way of filling the missing values.

- Fill in the missing value manually: This approach is time-consuming and may not be feasible given a large data set with many missing values.
- Use the attribute average value to fill in the missing value: This approach is not time-consuming and feasible given a large data set with many missing values.

In Data preprocessing module normalization the given data. In this paper, the data is normalized by min-max normalization [6]. The data were normalized in the range of (0.0, 1.0) by using as the normalizing factor the peak value. The data preprocessing output stored into database and then given to clustering module for clustering process.

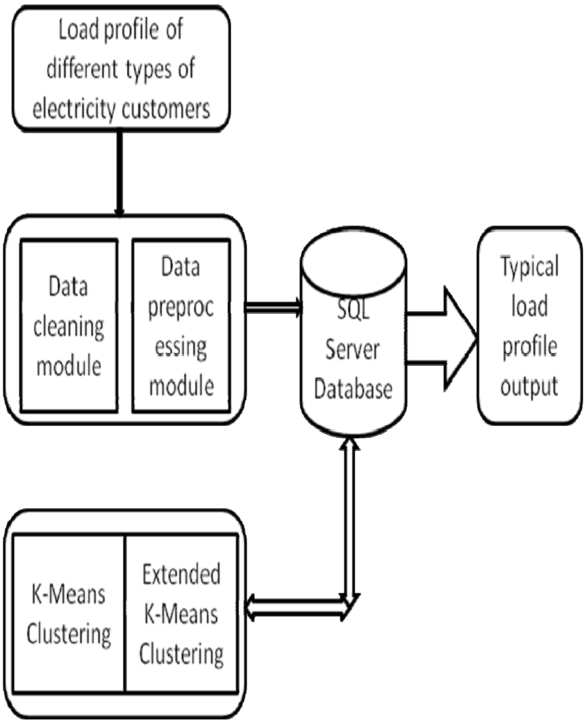

Fig.2 shows the architecture of system.

In clustering module two (k-means algorithm and extended k-means algorithm) algorithms used to form five clusters. Classical k-means clustering [8] groups a data set of $\mathbf{x}^{(n)}$ (n = 1, …, N) samples in k = 1, …, K clusters by means of an iterative procedure. A first guess is made for the K cluster centres $\mathbf{c}^{(k)}$ (usually chosen in a random fashion among the samples of the data set). The K centres classify the samples in the sense that the sample $\mathbf{x}^{(n)}$ belongs to cluster k if the distance $\| \mathbf{x}^{(n)} - \mathbf{c}^{(k)} \|$ is the minimum of all the K distances. The estimated centres are used to classify the samples into clusters (usually by Euclidean norm) and their values $\mathbf{c}^{(k)}$ are recalculated. The procedure is repeated until stabilisation of the cluster centres. Clearly, the optimal number of clusters is not known a priori and the clustering quality depends on the value of K. The clusters formed by k-means clustering algorithm contains the two different types of customers in same cluster.

The extended k-means algorithm helpful for calculating consumption of particular type of customer or particular region. The clusters which are formed by k-means algorithm is given to extended k-means algorithm. The extended k-means algorithm find maximum number of same type of customers from cluster and transfer the different type of customer to their respective cluster. The extended k-means algorithm useful in load scheduling problem.

3. RESULT AND DISCUSSION

Difficult to cluster the huge amount of electricity data. Figure 3 show the load profile of different types of customers (residential, commercial, industrial, street light etc.). Load profile of different types of customers normalized by min-max normalization method.

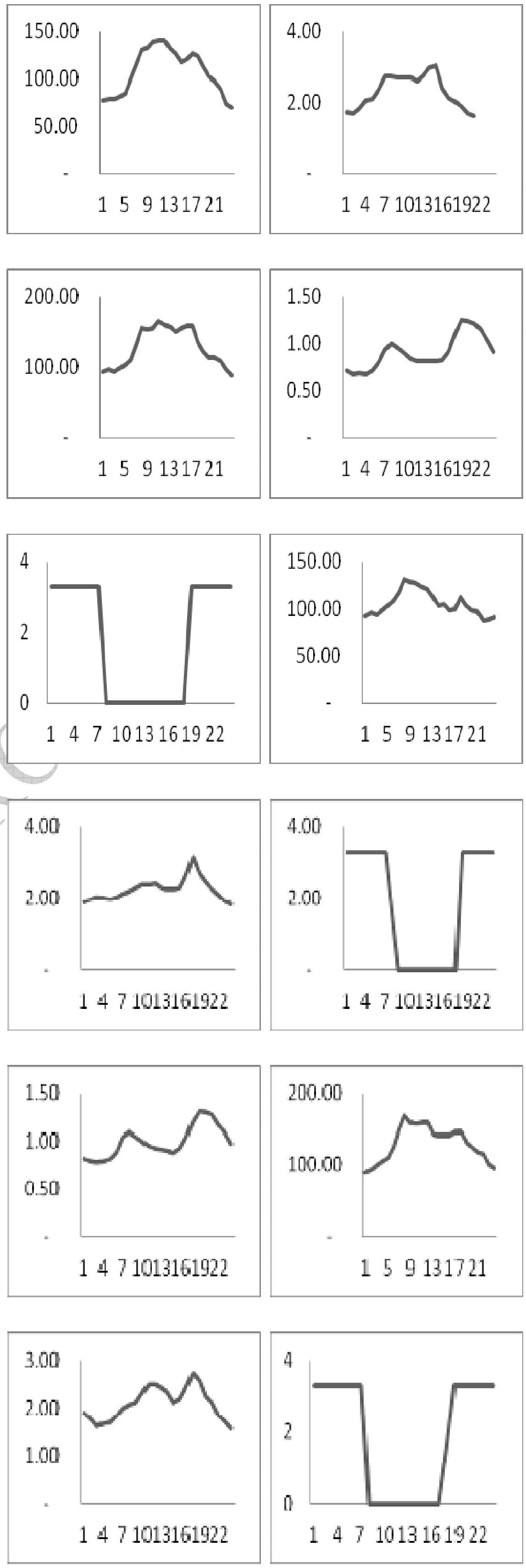

Fig.3 load curve of electricity customer (x-axis: time and y-axis: load in kw)

Hourly based data used for forecasting the load pattern of customers. Figure 4 show the cluster (k-means algorithm) result.

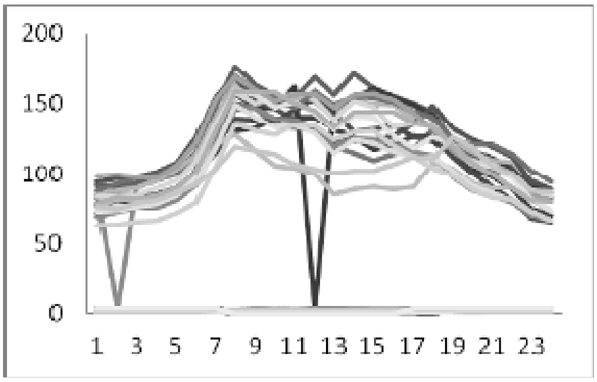

Fig.(a) cluster1

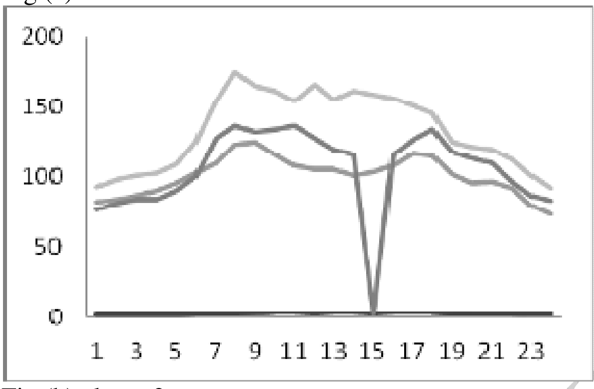

Fig.(b) cluster2

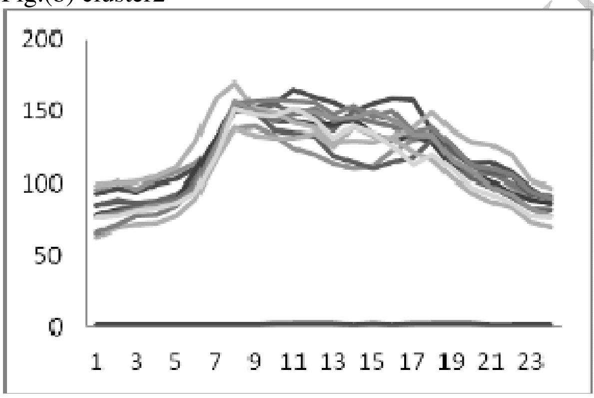

Fig.(c) cluster3

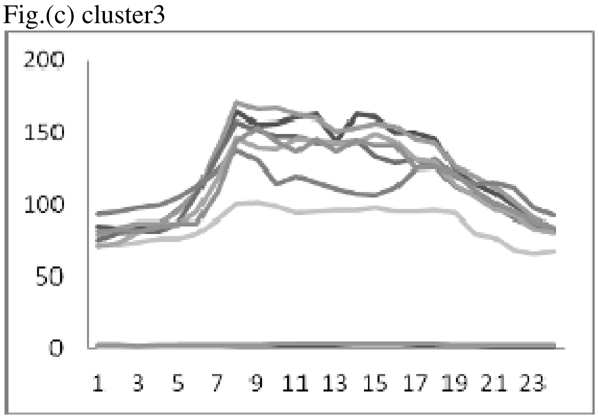

Fig.(d) cluster4

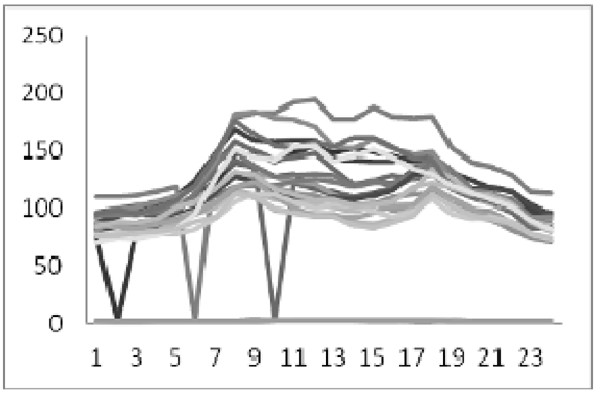

Fig.(e) cluster5

Fig.4(a-e) shows the clustering result of 5 clusters (x-axis: time and y-axis: load in kw)

After applying k-means algorithm the different types of customers clustered into 5 clusrters. Cluster1 contains the all types of customers, cluster2 contains three commercial and three industrial customers ,cluster3 contains one commercial and some industrial customers, cluster4 and cluster5 contains the commercial and industrial customers.

4. CONCLUSION

The paper presents the prediction of electricity of different types of customers. How to determine the typical load profile based on the clustering methods and generation typical load profile of different types of customers. Min-max normalization method scale the data into (0, 1). Different normalization methods may cause different clustering results. Paper shows the cluster result which helpful for forecasting the electricity and also helpful for deciding tariff plan for consumers.

In this work, same types of load pattern clustered by k-means algorithm. The extended k-means clustering helpful to form appropriate clusters. In k-means algorithm less number of cluster improve the classification rate.

5. REFERENCES

[1] A. H. Nizar , Z. Y. Dong , and Y. Wang, Power utility nontechnical loss analysis with extreme learning machine method, IEEE Trans. Power Syst., vol. 23, no. 3, pp. 946–955, Aug. 2008.

[2] G. Chicco and I. S. Ilie, Support vector clustering of electrical load pattern data, IEEE Trans. Power Syst., vol. 24, no. 3, pp. 1619–1628, Aug. 2009.

[3] G. Chicco, R. Napoli, and F. Piglione, Comparisons among clustering techniques for electricity customer classification, IEEE Trans. Power Syst., vol. 21, no. 2, pp. 933–940, May 2006.

[4] G. Chicco, R. Napoli, and F. Piglione, Application of clustering algorithms and self organising maps to classify electricity customers, in Proc. IEEE Bologna Power Tech Conf., Jun. 23–26, 2003, vol. 1.

[5] N. M. Kohan, M. P. Moghaddam, S. M. Bidaki, and G. R. Yousefi, Comparison of modified k-means and hierarchical algorithms in customers load curves

clustering for designing suitable tariffs in electricity market, in Proc. 43rd Int. Universities Power Engineering Conf., Padova, Italy, Sep. 1–4,2008, pp. 1–5.
[6] J. Han and M. Kamber, Data Mining: Concepts and Techniques, 2nd ed. San Francisco, CA: Morgan Kaufmann, 2006.
[7] Tiefeng Zhang, Guangquan Zhang, Jie Lu , Xiaopu eng, and Wanchun Yang, "A New Index and Classification Approach for Load Pattern Analysis of Large Electricity Customers", IEEE Trans. Power Syst., vol. 27, NO. 1, Feb. 2012.
[8] J. A. Hartigan and M. A. Wong, Algorithm AS 136: A k-means clustering algorithm, J. Roy. Statist. Soc. Series C (Appl. Statist.), vol. 28, no. 1, pp. 100–108, 1979.
[9] D. Gerbec, S. Gasperic, I. Smon, and F. Gubina, "Allocation of the load profiles to consumers using probabilistic neural networks," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 548–555, May 2005.
[10] W. Y. Li, J. Q. Zhou, X. F. Xiong, and J. P. Lu, "A statistic-fuzzy technique for clustering load curves," *IEEE Trans. Power Syst.*, vol. 22, no. 2, pp. 890–891, May 2007.
[11] F. Rodrigues, J. Duarte, V. Figueiredo, Z. Vale, and M. Cordeiro, A comparative analysis of clustering algorithms applied to load profiling, Mach. Learn. Data Mining Pattern Recognit., Lecture Notes Comput. Sci., vol. 2734, no. 2003, pp. 73–85, 2003.
[12] S. C. Cerchiari, A. Teurya, J. O. P. Pinto, G. Lambert-Torres, L. Sauer, and E. H. Zorzate, Data mining in distribution consumer database using rough sets and self-organizing maps, in Proc. IEEE/PES Power Systems Conf. Expo., Atlanta, GA, Oct. 29–Nov. 1, 2006, pp. 38–43.
[13] D. Hand, H. Manilla, and P. Smyth, Principles of Data Mining. Cambridge, MA: MIT Press, 2001.