Clustering of Electricity Demand to Generate Virtual Load Profile

Marliah Binti Mostakim

Faculty of Electrical Engineering
Universiti Teknologi MARA Malaysia
40450 Shah Alam, Selangor, Malaysia
e-mail: marliah_mustakim@yahoo.com.my

Abstract—**Recently the emerging issue in the electric industry is effective power based on Smart Grid. To operate the power effectively, the data must be applicable and accessible, thus will produce the virtual load profile (VLP). To generate VLP clustering and classification are required. The clustering of customers electricity demand becomes important not only to design tariff but also to identify sets of standard load profile. Electricity demand means the maximum amount of electricity is being used at some time while the load profile can refer to a number of different forms of data. Clustering is one of the methods that can be used to perform the data. Clustering represent groups of customers with the same clusters are very similar and the different clusters become very distinct. In this paper, focus is on K-means and Hierarchical for clustering electricity demand and their differences are analyzed.**

Keywords-Clustering, Electricity Demand, Hierarchical, K-mean, Virtual Load Profile, Smart Grid.

I. INTRODUCTION

Smart Grid is a new and intelligent power system that has wide advantage for electrical power industry [1]. With Smart Grid the real time pricing will be practical and can operate efficiently. However the implementation is more costly. Facing this reality, therefore load profiling seems the alternative solution that would provide cost-effective approach since the efficient method; the direct monitoring was required cost-prohibitive by installing time intervals meters [2]. Load profiles have been used to provide important information to support multiple functions of electric utilities for system planning and operation [3]. To perform more details of load profile, actual demand can be collected at strategic location and it can provide benefits to look for peak consumption. Load profile varied according to customers' type and will perform the graph of electrical load versus time.

By knowing customer's load profile, it can simply determine the price of the customer's demand. Thus they can provide better marketing strategies and improve efficiency [4]. Electricity demands in Malaysia have been increasing due to development in the industrial and commercial sector. By performing the clustering analysis, objects within one cluster can share more in common with another cluster.

Cluster analysis is a way to examine similarities and dissimilarities of observations or objects. Data often fall naturally into groups, or clusters, of observations, where the characteristics of objects in the same cluster are similar and the characteristics of objects in different clusters are dissimilar.

Figure 1 (a) and 1 (b) illustrate the daily customers' electricity demand for complete 24 hours electricity consumption. This daily load curve contains 246 feeders and the measurement was monitored every half an hour interval. The graph of electricity demand (in p.u) versus time (for every half an hour) was performed.

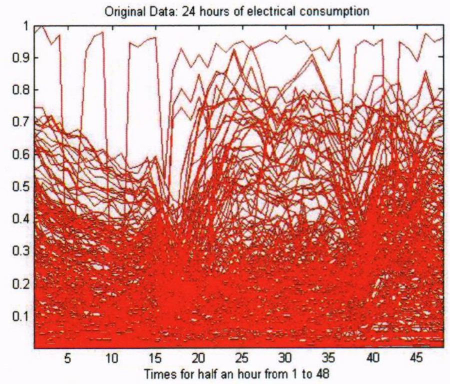

Figure 1 (a): Original Data for Customers Electricity Demand

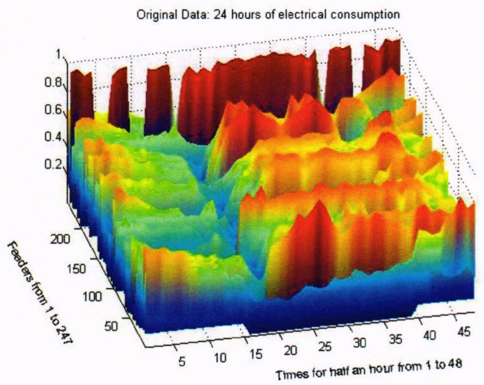

Figure 1 (b): Customers Electricity Demand in Surface View

The task of ensuring that virtual load profile (VLP) is derived from half-hour pattern must be completed before offering value-add smart grid services [5].

In this paper, the ability of Hierarchical and K-means Clustering to generate the VLP is analyzed and their performances are examined.

The remaining parts are summarized as follows: Section II discussed briefly on clustering techniques; Section III study the result for their performance; and lastly Section IV contains the conclusion.

II. CLUSTERING TECHNIQUE

Clustering techniques are performed to classify electricity consumers' demand. In the literature, various clustering techniques have been identified [1-12]. This section will review some several techniques with their procedures.

A. K-means Clustering

K-means Clustering is a partitioning method. The function ***kmeans*** partitions data into k mutually exclusive clusters, and returns the index of the cluster to which it has assigned each observation.

For the starting, K-means initialize the centroid and the desired number of clusters. Then it can create the silhouette plot. Next, increase the number of clusters to see if K-means can find a better grouping of the data. A more quantitative way to compare the two solutions is to look at the average silhouette values. When it meet the requirements it will end, if not we can do again from the step that increase the number of cluster. These procedures are illustrated at Figure 2 (a):

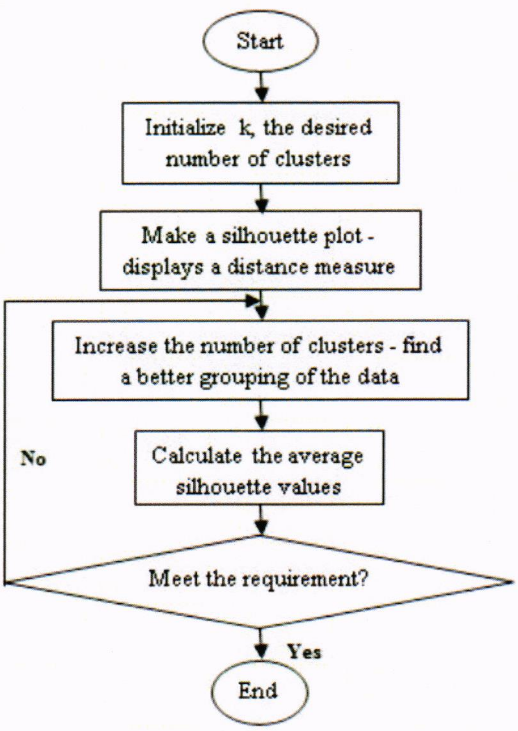

Figure 2 (a): Procedure for K-means clustering

The silhouette plot displays a measure of the distance in one cluster to another cluster. This measure ranges from +1, that are very distant from other clusters, through 0, indicate not distinctly, to -1, probably wrong cluster.

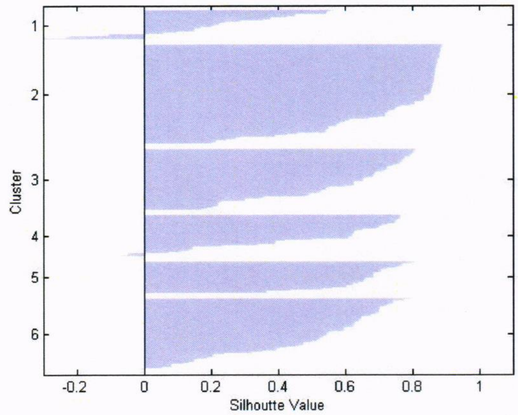

Figure 2 (b): Silhoutte for K-means Clustering

The distinctions mean that k-means clustering is often more suitable than hierarchical clustering for large amounts of data. Over all clusters, K-means uses an iterative algorithm which minimizes the sum of distance from each load profile to its cluster centroid load profile [6].

B. *Hierarchical Clustering*

Hierarchical Clustering groups data over a variety of scales by creating a cluster tree or dendrogram. As mentioned in [4-6], the tree is not a single set of clusters, but rather a multilevel hierarchy, where clusters at one level are joined as clusters at the next level.

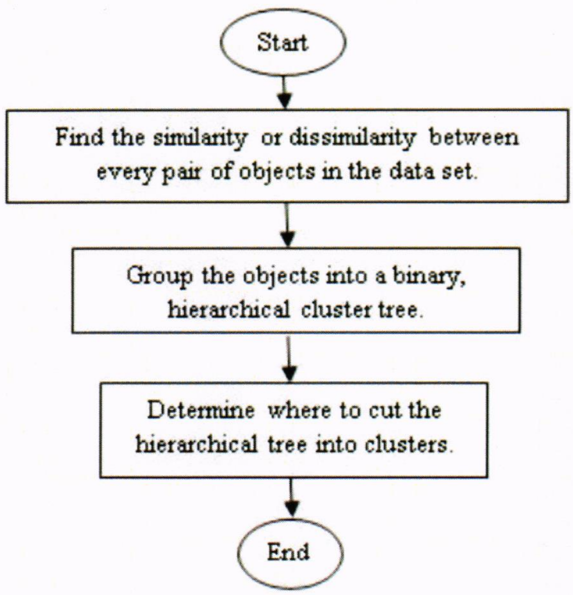

Figure 2 (c): Procedure for Hierarchical Clustering

Figure 2 (c) reveals the procedure for hierarchical clustering. First of all is to find the similarity or dissimilarity between every pair of objects in the data set. In this step, it will calculate the distance between objects using the ***pdist*** function. Then, group the objects into a hierarchical cluster tree. As objects are paired into binary clusters, the newly formed clusters are grouped into larger clusters until a hierarchical tree is formed. Last, determine where to cut the hierarchical tree into clusters. This creates a partition of the data.

Figure 2 (d) and 2 (e) display the dendrogram form where the x-axis represents the feeder data sets while the y-axis represents the distance between clusters. Figure 2 (d) represent all the feeders involve while Figure 2 (e) reveals when compute after ***pdist*** and ***linkage.***

Figure 2 (d): Dendrogram for Overall Feeder

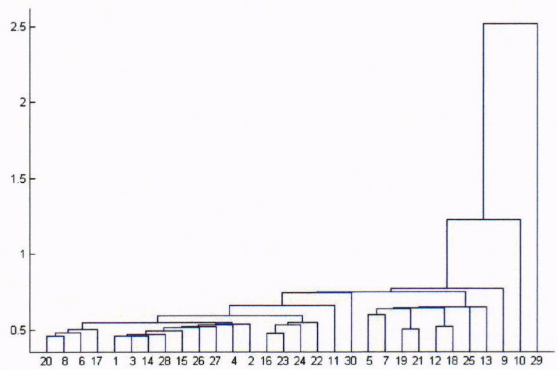

Figure 2 (e): Dendrogram for Hierarchical Clustering

III. Results And Discussion

Both clustering were tested to set load data for different feeders. This daily load curve contains 246 feeders and the measurement was monitored every half an hour interval. As referred in [2], the data is taken from 12 midnight until 11.30pm the following day. Therefore the load profile is conducted by 48 load values throughout the day.

A. K-means Clustering

K-means clustering is an unsupervised clustering algorithm. "*K*" stands for the number of clusters. Some criteria can be used to automatically estimate *K*. *K*-means algorithm is iterative. Only a local minimum is obtained although it converges. *K*-means works only for numerical data and easy to implement.

To reveal how well-separated the resulting clusters are, plotting silhouette of the cluster is done by using the output from K-means. These silhouette plots illustrate at Figure 3 (a), 3 (b) and 3 (c). Average silhouette values are 0.5611, 0.6637 and 0.5749. Since the silhouette plot for three clusters are better separated than the cluster 2 and 4, therefore 0.6637 is chosen as the best silhouette value.

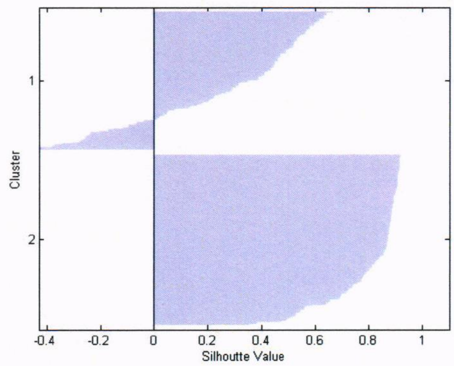

Figure 3 (a): 2 clusters – average 0.5611

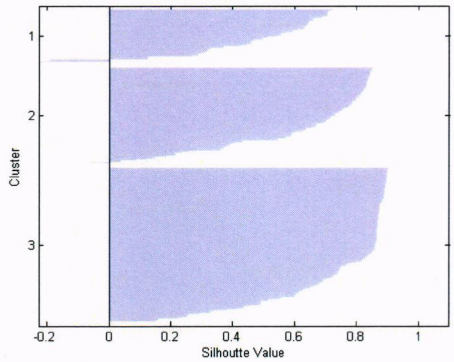

Figure 3 (b): 3 clusters – average 0.6637

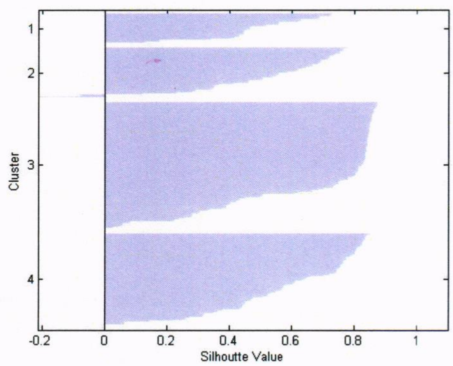

Figure 3 (c): 4 clusters – average 0.5749

Based on the silhouette plot, 0.6637 is found as a suitable value because it gives well-separated results compares to others. Using this value, the data was classified into three clusters and the cluster then plotted based on silhouette.

$$\boxed{A = \text{mdwtcluster(silh)}}$$

The entire graphs plotted are summarized below. The horizontal part shows the hourly time and the vertical part shows power in p.u value.

Figure 3 (d) display the domestic feeder since it decreases in the morning and then increases in the evening. Figure 3 (e) reveals the commercial feeder which is early in the morning it increases until reach the maximum point before it slowly decreases in the evening.

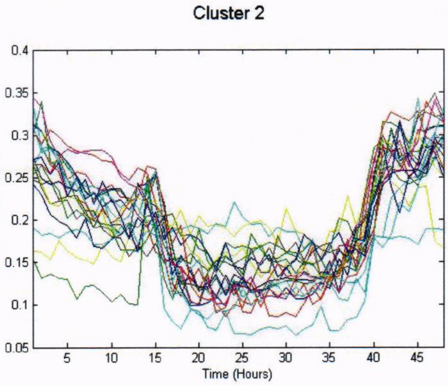

Figure 3 (d): Cluster 2 (K-means)

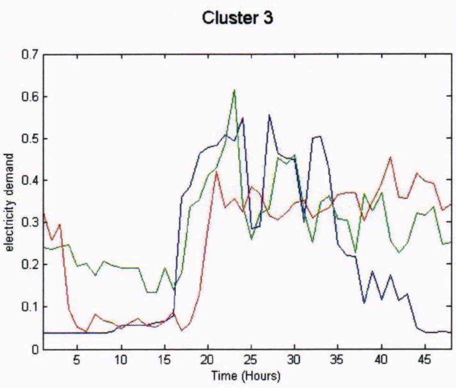

Figure 3 (e): Cluster 3 (K-means)

Combination feeder has shown in the Figure 3 (f). It is between commercial feeder, domestic feeder and industrial feeder. Industrial feeder is operating nonstop.

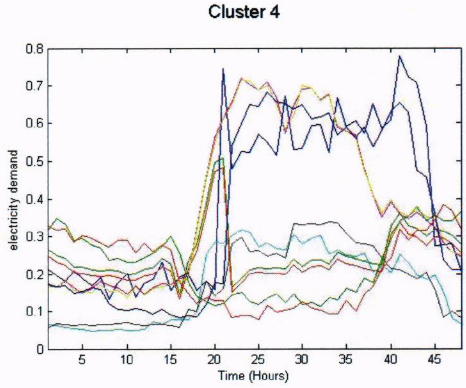

Figure 3 (f): Cluster 4 (K-means)

B. Hierarchical Clustering

With Hierarchical clustering, a dendrogram tree is created regardless of the final number of cluster. Then the number of clusters needed is taken into account, and the cutting position is determined [4-6].

Similarly, the number of cluster must be specified before performing Hierarchical analysis. The hierarchical cluster tree may naturally divide the data into well-separated data. But since the optimal number of cluster is three in K-means, therefore the same number of cluster is used here. This is to ensure the differentiation of analysis is easier.

The result below is when plot overall feeder:

$$\boxed{S = \text{mdwtcluster(X)}}$$

S is constructs clusters from a hierarchical cluster tree. The input matrix X is decomposed in row direction with the maximum allowed level.

The load curves in each cluster are portrays in Figure 3 (g) to (i). Figure 3 (g) shows the commercial feeder since it increases in the morning and decreases in the evening.

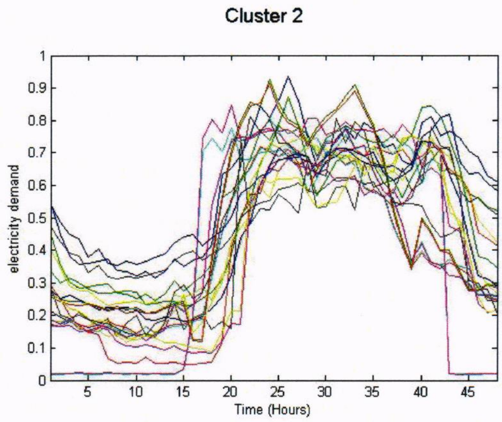

Figure 3 (g): Cluster 2 (Hierarchical)

For Figure 3 (h) display the domestic feeder, which in the morning the load pattern is decreased while in the evening rapidly is increased. The combination feeder illustrated in Figure 3 (i) consists of commercial, domestic and industrial feeder.

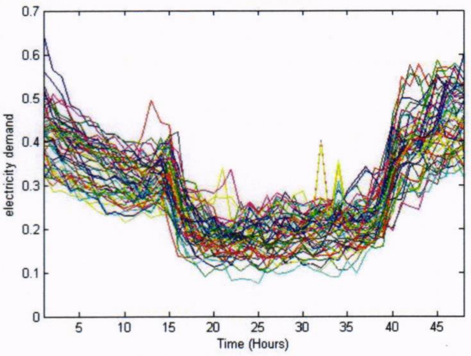

Figure 3 (h): Cluster 3 (Hierarchical)

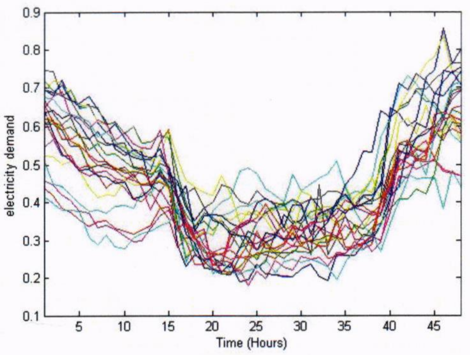

Figure 3 (i): Cluster 4 (Hierarchical)

C. *Analysis from both approaches*

Since there were too many feeders involved, only three clusters can delegate to be presented. This is based on the silhouette plot that gives the best average value.

In K-means the domestic feeder can display in the cluster 2 and commercial feeder in cluster 3. Since the algorithm is different, these are not necessarily the same clusters as those found by Hierarchical clustering. So in Hierarchical, it shown opposite from K-means. Cluster 2 is commercial and cluster 3 is domestic feeder.

Both clustering show the combination feeder in cluster 4. The results do not illustrate very well, but if we see throughly, the combination of the three feeder can be seen. This may because the three feeder operate similarly but in different loads. Thus in K-means, commercial suppresed the industrial pattern while in Hierarchical the domestic feeder dominate the whole pattern.

IV. CONCLUSION

This study gives the various techniques for clustering electricity demand. According to the analysis, each technique has its own characteristics. Hierarchical gives the easy way to group the data by analyzing the dendrogram. K-means operates on actual observations, if the number of the cluster has been decided.

From the cluster obtained, Hierarchical clustering has reveals their ability to generate virtual load profile than K-means clustering. Further investigation is necessary to classify the electricity demand to generate virtual load profile.

ACKNOWLEDGMENT

I would like to take this opportunity to show my gratitude to my research advisor, Assoc. Prof. Dr. Zuhaina Hj Zakaria for her advices to help me for the whole process of this report. I feel blessed to have her as my advisor. I thank her for every piece of her intensive efforts that have been put into this research work.

REFERENCES

[1] Khajavi, P.; Monsef, H.; Abniki, H.; , "Load profile reformation through demand response programs using Smart Grid," Modern Electric Power Systems (MEPS), 2010 Proceedings of the International Symposium , vol., no., pp.1-6, 20-22 Sept. 2010

[2] Z. Zakaria, M. N. Othman, and M. H. Sohod, "Consumer Load Profiling using Fuzzy Clustering and Statistical Approach," in Research and Development, 2006. SCOReD 2006. 4th Student Conference on, 2006, pp. 270-274.

[3] Zuhaina Hj Zakaria; Lo, K.L.; , "Load profiling in the new electricity market," Research and Development, 2002. SCOReD 2002. Student Conference on , vol., no., pp. 278- 281, 2002

[4] Gerbec, D.; Gasperic, S.; Smon, I.; Gubina, F.; , "An approach to customers daily load profile determination," Power Engineering Society Summer Meeting, 2002 IEEE , vol.1, no., pp.587-591 vol.1, 25-25 July 2002

[5] Young-Il Kim; Jong-Min Ko; Seung-Hwan Choi; , "Methods for generating TLPs (typical load profiles) for smart grid-based energy programs," Computational Intelligence Applications In Smart Grid (CIASG), 2011 IEEE Symposium on , vol., no., pp.1-6, 11-15 April 2011

[6] Young-Il Kim; Shin-Jae Kang; Jong-Min Ko; Seung-Hwan Choi; , "A study for clustering method to generate Typical Load Profiles for Smart Grid," Power Electronics and ECCE Asia (ICPE & ECCE), 2011 IEEE 8th International Conference on , vol., no., pp.1102-1109, May 30 2011-June 3 2011

[7] Prahastono, I.; King, D.; Ozveren, C.S.; , "A review of Electricity Load Profile Classification methods" Universities Power Engineering Conference, 2007. UPEC 2007. 42nd International , vol., no., pp.1187-1191, 4-6 Sept. 2007

[8] Prahastono, I.; King, D.J.; Ozveren, C.S.; Bradley, D.; , "Electricity load profile classification using Fuzzy C-Means method," Universities Power Engineering Conference, 2008. UPEC 2008. 43rd International , vol., no., pp.1-5, 1-4 Sept. 2008

[9] In Hyeob Yu; Jin Ki Lee; Jong Min Ko; Sun Ic Kim; , "A method for classification of electricity demands using load profile data," Computer and Information Science, 2005. Fourth Annual ACIS International Conference on , vol., no., pp. 164- 168, 2005

[10] Meffe, Andre; Oliveira, Carlos; , "Classification techniques applied to electrical energy distribution systems," Electricity Distribution, 2005. CIRED 2005. 18th International Conference and Exhibition on , vol., no., pp.1-5, 6-9 June 2005

[11] Azevedo, F.; Vale, Z.A.; , "Forecasting Electricity Prices with Historical Statistical Information using Neural Networks and Clustering Techniques," Power Systems Conference and Exposition, 2006. PSCE '06. 2006 IEEE PES , vol., no., pp.44-50, Oct. 29 2006-Nov. 1 2006

[12] Gerbec, D.; Gasperic, S.; Smon, I.; Gubina, F.;, "Consumers' Load Profile Determination Based on Different Classification Methods," Faculty of Electrical Engineerin, 2003 IEEE.