Methods for Generating TLPs (Typical Load Profiles) for Smart Grid-Based Energy Programs

Young-Il Kim
S/W Center
KEPCO Research Institute
Daejeon, Korea
yikim@kepco.co.kr

Jong-Min Ko
S/W Center
KEPCO Research Institute
Daejeon, Korea
kojm@kepco.co.kr

Seung-Hwan Choi
S/W Center
KEPCO Research Institute
Daejeon, Korea
captain@kepco.co.kr

***Abstract*— Most electric power companies implement an automatic meter reading (AMR) system operating at quarter-hour intervals. The companies install the system (an electric meter) at the homes of their high-voltage (HV) customers who consume a significant amount of power each month. When first introduced to the industry, the AMR system simply measured customers' peak power and billed them. Recent studies are examining the same system's applicability in the cutting-edge smart grid technology. A growing number of studies are focusing on AMR-based distribution network load analysis and demand prediction to promote the dissemination of smart grid-based services. Researchers are basically using AMR customers' usage data to analyze loads and generate the virtual load profile (VLP) of non-automatic meter reading (nAMR) customers. Generating VLP requires clustering and classification that are among the various data mining techniques adopted by researchers. This study reviewed previous research findings that reported AMR-based typical load profile (TLP) generation, and utilized the AMR data of some KEPCO HV customers for TLP generation. Analyses were performed via three clustering techniques, and the strengths of the techniques were compared.**

Keywords-load profile; smart grid; typical; virtual

I. INTRODUCTION

A growing number of countries are integrating or combining cutting-edge communication, control, electronics, and information technologies into the energy industry to upgrade their current power grid to the smart grid system. The upgrade involves improvements and solidification to power grid security, quality, reliability, and availability, as well as power transmission, substation system, power supply, and smart consumption. The focal point of energy research has been in smart grid architecture [1]. Smart grid based on advanced telecommunications and computing techniques would influence restructured electricity industry [2]. It offers communication, control, and services that are capable of realizing dynamic energy management by integrating technical elements such as energy efficiency, demand response, and AMI (advanced metering infrastructure).

To offer high-value power services via the smart grid system, it is imperative first to identify consumers' load patterns through systems currently available in the market, which include energy management and metering systems and related information systems. Next, power companies must be able to provide information and data such as load analysis data for power distribution systems and consumer demand predictions. An even more important consideration is that energy companies need to manage quarter-hour power usage by their customers and facilities to offer full-scale smart grid-based services to the market. Realistically, however, implementing this approach remains a tall order due to technology and cost issues. Researchers, therefore, are focusing on the creation of virtual quarter-hour data by processing manual metering-based monthly consumption data that can be obtained from existing automatic metering-based quarter-hour data.

Power (kWh) that is supplied through a specific transformer in the distribution system corresponds to the sum of power (kWh) consumed by households serviced by that transformer, unless the loss that occurs during distribution is taken into account and deducted accordingly. Thus, knowing the quarter-hour power consumption of all of the customers connected through the transformer will allow power companies to estimate the transformer's average quarter-hour supply of power without physically installing the quarter-hour metering system in the transformer. This, however, presents yet another challenge that the companies have no choice but to install the meter at each home serviced through the transformer. The most cost-effective and realistic alternative — and the increasingly more popular research theme — would be then to use the load patterns of customers equipped with AMR (automatic meter reading) system at home and to apply them in calculating the virtual quarter-hour power (kWh) of nAMR (non-automatic meter reading) customers. This application makes it possible to obtain nAMR customers' virtual load curve and AMR customers' load curve. Applying the curves to power facilities' GIS (geographic information system) data and analyzing them allows energy companies to carry out load analysis of the distribution network circuit as well as the facilities including distribution zones, transformers, and banks. Time data mining and spatial data mining techniques will also make it possible to estimate power consumption patterns.

This paper is organized as follows: (a) an introduction (Chapter 1); (b) methods for creating virtual load patterns by using AMR data and by generating typical load patterns (Chapter 2); (c) review of previous research findings in typical load patterns and the experimental methods adopted in this paper (Chapter 3); (d) the results obtained from typical load pattern creation and performance analysis, which were made possible by applying the data of HV (high-voltage) customers of KEPCO (Korea Electric Power Corporation) (Chapter 4); and (e) conclusions (Chapter 5).

II. Typical Load Profile and Virtual Load Profile

The most urgent task to be completed before offering value-added smart grid services to customers is to ensure a virtual load profile (VLP) that is derived from quarter-hour power consumption patterns of nAMR consumers. The basic principle in creating VLP is matching nAMR customers' daily power consumption to daily load patterns of consumers who exhibit similar consumption patterns. Take two large supermarkets, Mart A (1,000m2) and Mart B (500m2). Mart A is an AMR customer and Mart B is a nAMR customer. Their monthly usage is 10,000 kWh (Mart A) and 5,000 kWh (Mart B). The perfectly proportionate relationship of the two markets' area and usage indicate that they are likely to have similar consumption patterns. Taking this assumption a step further, energy companies could generate the virtual load curve of Mart B by reducing Mart A's actual load curve by 50%.

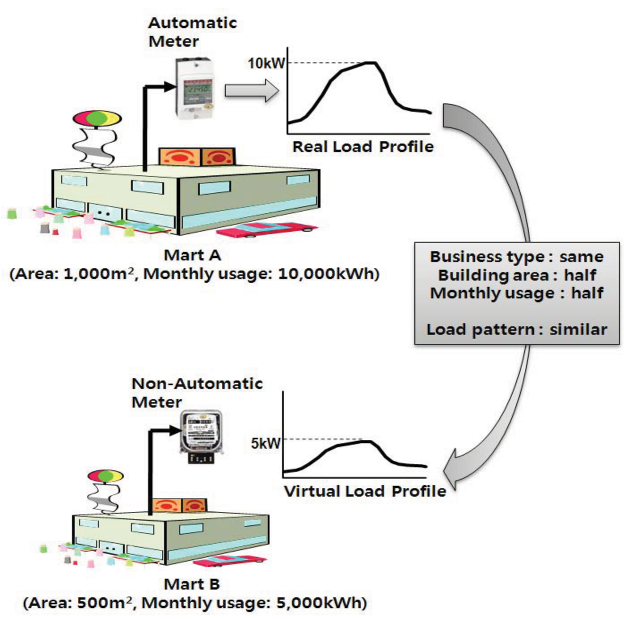

Figure 1. Conceptual model of virtual load profile generation

To apply the conceptual model of VLP generation to all nAMR customers, one must implement a clustering process in which researchers group AMR customers according to similar load patterns, and then assign a typical load profile (TLP) for each group. In addition, researchers must classify nAMR customers into sub-groups that will match those of AMR customers with similar usage patterns.

Once the clustering and classification processes are implemented, the following phases need to be followed to generate the VLP of nAMR customers:

- Carry out pre-processing and normalization on the AMR data that were collected;
- Generate clusters using the normalized AMR data, and determine a TLP for each group;
- Create classification rules based on which highly associated groups will be identified by their common attributes;
- Find the AMR groups that each match nAMR groups by applying nAMR groups' common attributes to the classification rules, and select TLP's for the matching groups; and
- Apply nAMR customers' monthly usage to the selected TLP's and generate the VLP.

The first phase listed above is normalization. At this stage, energy companies extract service program-specific data and AMR data of their customers to create TLP's. They implement pre-processing to eliminate inaccurate quarter-hour data resulting from imperfect metering resulting from communication failure, etc. The pre-processed data are normalized for each customer so that the sum of customers' quarter-hourly load becomes '1.'

The second phase is clustering. It involves grouping customers with similar load patterns into several clusters. Figure 2 illustrates this concept where actual load patterns of AMR customers (A1-A6) are used as experimental data for clustering. Using the clustering algorithm that is frequently applied in data mining, the load patterns (A1 - A6) can be organized into several clusters featuring similar profiles. At this point, the mean or representative value of the actual load patterns of the customers in each cluster is calculated and defined as the TLP for that cluster. The following formulas show how the C1 cluster's quarter-hour TLP was calculated by using the daily load pattern of customer groups A5 (L_{day}^{A5}) and A6 (L_{day}^{A6}), both of which belong to C1 as shown in Figure 2.

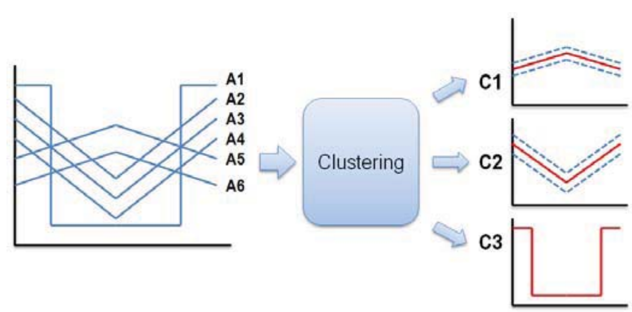

Figure 2. Clustering phase

$$L_{day}^{A5} = [\, l_1^{A5}, l_2^{A5}, \dots, l_i^{A5}, \dots, l_{96}^{A5}]$$

$$L_{day}^{A6} = [\, l_1^{A6}, l_2^{A6}, \dots, l_i^{A6}, \dots, l_{96}^{A6}]$$

$$TLP_{day}^{C1} = \left[\frac{l_1^{A5} + l_1^{A6}}{2}, \dots, \frac{l_i^{A5} + l_i^{A6}}{2}, \dots, \frac{l_{96}^{A5} + l_{96}^{A6}}{2}\right]$$

The third phase is classification. The attributes of AMR customers and their cluster numbers are utilized to make up the classification rules, which will automatically assign a certain AMR customer group to its matching cluster. Once the rules are established, all there is left to do is use the attributes of nAMR customers to determine which AMR groups they each correspond with, and to which cluster they belong.

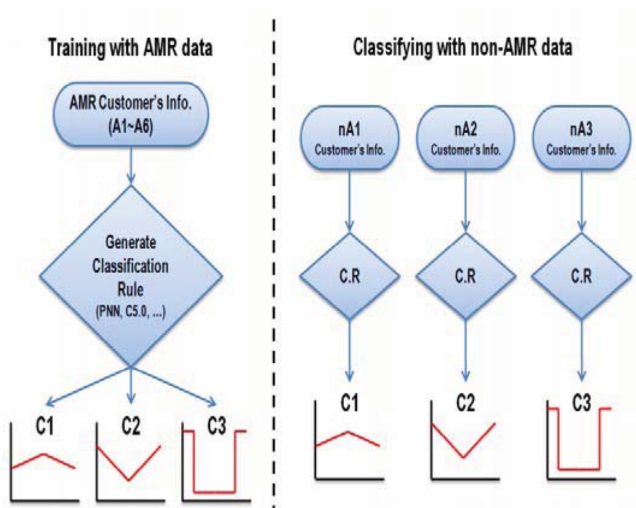

Figure 3. Classification phase

The fourth and fifth phases correspond with VLP generation. The daily VLP of nAMR customers is created by using the customers' monthly usage and TLP's. For instance, nA1 (nAMR customer) uses 3,000 kWh of power each month. Translated into daily usage, 100 KWh of power is consumed each day. The figure will then be applied to a TLP curve that was typified to have the total daily usage of '1' to calculate the total monthly usage of 100 kWh. That way, the VLP is generated.

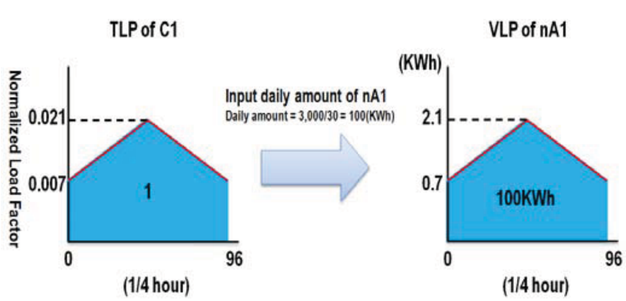

Figure 4. Virtual load profile generation phase

III. Experimental Methods

A. TLP Generation Methods

Generally speaking, energy companies cluster their customers and generate TLP based on decades-long industry practices and experience as well as the attributes and annual usage of their customers. However, the conventional methods do not estimate accurately the load patterns of customers, and thus they contribute to a number of errors made during load analysis, demand prediction, and other related processes. More recently, researchers have started to use other intelligent ways to estimate the loads, such as an artificial neural network (ANN) [3], support vector machines (SVM's), and a self-organizing map (SOM). In addition, analysis of customers' load patterns is being conducted via k-means, SOM, fuzzy c-means (FCM), and hierarchical clustering [4].

Teemu Rasanen [5] suggested a technique that utilizes customers' load patterns to ensure more efficient energy operations and consumer-oriented services. The researcher used SOM and k-means algorithms to conduct analysis of the bulk, time-series power usage data. For the study, 3,989 LV (low-voltage) customers in Finland's Northern-Savo Province had participated. Their hourly usage data were used to group them into clusters and to calculate the mean load curve.

A study by Iswan Prahastone [6], on the other hand, reported efforts to create a number of different customer classification systems, which aimed at establishing a variety of billing programs. Techniques used for the study include hierarchical clustering; k-means clustering; FCM clustering; follow-the-leader clustering; and fuzzy relation clustering. Table 1 compares these clustering techniques. Hierarchical clustering, for instance, is capable of generating tree or dendrogram without having to determine the number of clusters in advance. Follow the leader and fuzzy relation clustering methods, on the other hand, offer useful algorithms that can be used when bulk fuzzy data are available.

TABLE I. Comparison of Clustering Methods

Methods	Number of clusters is predetermined	Creates boundaries between data sets	Requires Iterative process	Trial and error approach
Hierarchical	No	Yes	No	No
K-means	Yes	Yes	Yes	No
Fuzzy c-means	Yes	No	Yes	No
Follow the leader	If necessary	Yes	Yes	Yes
Fuzzy relation	If necessary	No	Yes	Yes

B. Experimental Methods

Analysis of the previously described clustering techniques found that hierarchical, k-means, and FCM were most appropriate for experimentation since they do not require fuzzy data.

1) *Hierarchical Clustering*

Hierarchical clustering groups data over a variety of scales by creating a cluster tree or dendrogram. 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. Hierarchical clustering is carried out in the following steps [6]:

- Similarity is determined by pairing load-pattern data sets;
- Define two similar data sets as a group, and create a cluster tree with inheritance profiles by using the limit values of the similarity; and
- Determine a cutting position that is capable of suspending the hierarchical cluster tree.

Figure 5 illustrates a dendrogram, in which a hierarchical cluster tree is used to cluster load profiles data sets. The horizontal axis refers to load profiles data sets, while the vertical axis reflects the distance between clusters. The dotted line indicates the cutting position of the cluster tree. With hierarchical clustering, deciding the cutting position determines the total number of clusters.

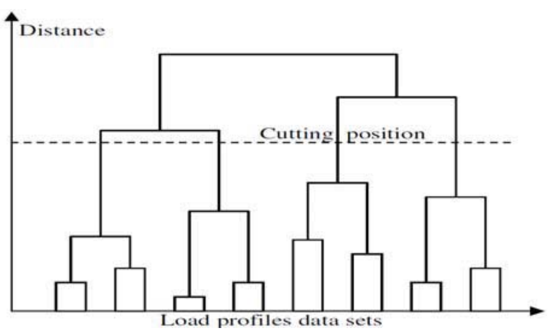

Figure 5. Dendrogram of hierarchical clustering

2) *K-means Clustering*

The K-means algorithm is one of the clustering methods combining adjacent data in a specific area and dividing them into several groups. Assuming that data is a dot of vector space, it minimizes the degree of dispersion of groups by classifying the dots in each group to minimize the Euclidean distance to the group where a customer belongs compared to the Euclidean distance to the central value of other groups. It can be summarized as follows:

$$V = \sum_{i=1}^{k} \sum_{j \in S_i} \left| x_j - \mu_i \right|^2, \text{ where,}$$

S_i: The i-th clusters, i={1,2,…,k}

μ_i : Central(typical) value of each group

First, the k-means algorithm generates the Ci group by dividing each dot into k groups and attains central value μi of each Ci. A new Ci is then created by connecting each dot to the nearest central value among the attained values. After repeating it, if the dots in each group are no longer moved to the other group or are converted into a status wherein the central value is not changed, the central value at that point is determined as a typical value of the related group.

3) *Fuzzy C-means Clustering*

The Fuzzy c-means clustering [7] is very similar to the k-means clustering. The fuzzy c-means is a data clustering technique wherein each data point belongs to a cluster to some degree that is specified by a membership grade. The degrees of membership for each data set to all clusters should sum to 1. The algorithm minimizes intra-cluster variance as well, but has the same problems as k-means; the minimum is a local minimum, and the results depend on the initial choice of weights.

IV. EXPERIMENTAL RESULT

This study sought to generate the typical load profile (TLP) to be used by KEPCO, and identify which of the conventional clustering algorithms is most suitable for generating the intended TLP. The study had 3,183 HV customers whose AMR data were used for the generation. To conduct performance analysis, hierarchical clustering, k-means clustering, and FCM clustering algorithms were used, and TLP was generated with clusters that ranged from 2 to 14 in number. Also, this experiment calculated the processing time of the algorithms; and the MAE (mean absolute error) from the customers' actual load profiles and the TLP.

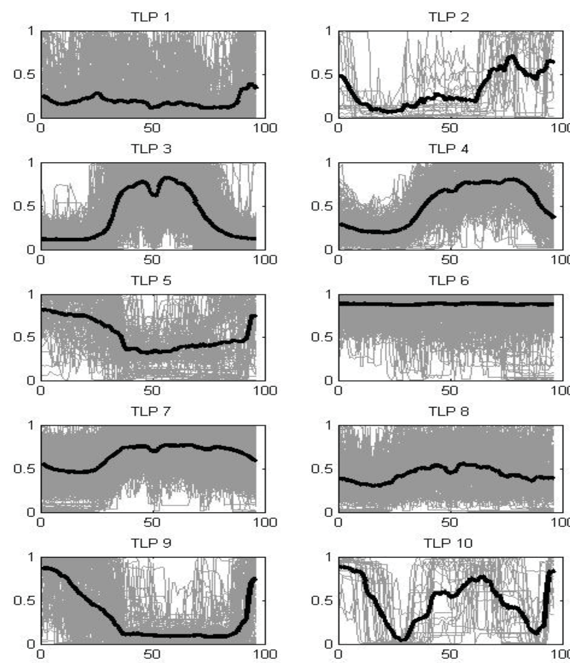

Figure 6. TLP and RLP of hierarchical clustering

Figure 6 shows the TLP (thick line), which was generated by grouping the participants into 10 clusters via hierarchical clustering, and the real load profiles (RLP) (gray line) of the customers in each cluster. Figure 7 lists the TLP and RLP obtained via k-means clustering, and Figure 8 the same data obtained via fuzzy c-means clustering. Reviewing each of the three figures (Figures 6 - 8) reveals that with hierarchical clustering, the similarity between TLP's has been minimized, which resulted in fewer similarly-configured TLP's in the graphs. In contrast, the graphs indicate the k-means clustering technique resulted in similarity between TLP 2, TLP 6, TLP 8, and TLP 9. With fuzzy c-means clustering, the graphs show similarity between TLP 2, TLP 3, TLP 5, TLP 8, and TLP 10.

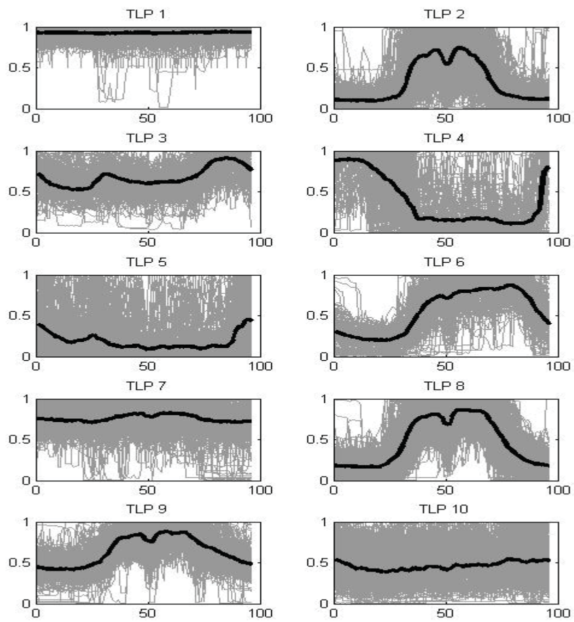

Figure 7. TLP and RLP of k-means clustering

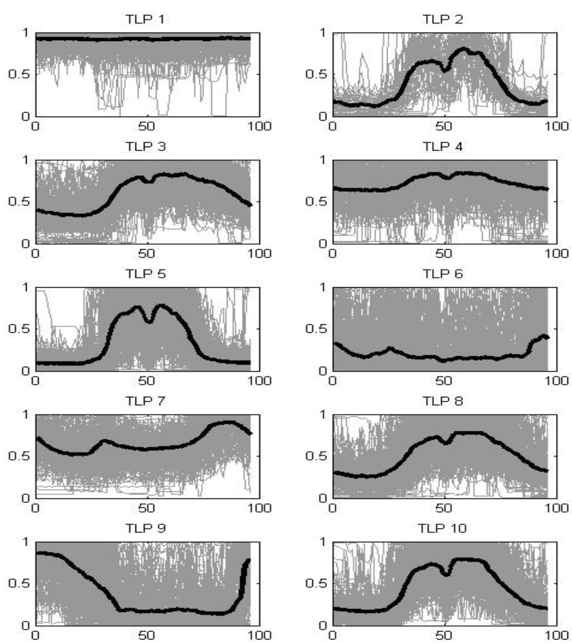

Figure 8. TLP and RLP of fuzzy c-means clustering

Figure 9 illustrates the processing time that is required to generate TLP in relation to the cluster count. The processing time was calculated by using the three clustering techniques; hierarchical, k-means, and fuzzy c-means, respectively. With hierarchical clustering, a cluster tree is created regardless of the final number of clusters. Then, the number of clusters needed is taken into account, and the cutting position is determined. The processing, for the most part, involves the generation of the tree, and thus processing time is consistent regardless of the cluster count. With k-means clustering, a shorter processing time is generated than hierarchical clustering when the number of clusters is small, but increases as the cluster counts grow larger. Fuzzy c-means clustering involves not only the processing that calculates the mean of clusters but another iteration as well to calculate the membership value for each data set. Thus, fuzzy c-means clustering requires a minimum of twice the processing time needed for k-means clustering.

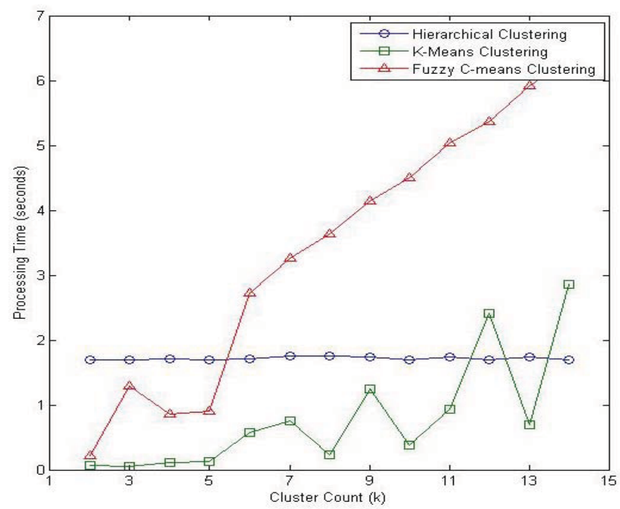

Figure 9. Processing time in relation to cluster count

Figure 10 lists graphs that indicate the MAE (mean absolute error) between each cluster's TLP and the cluster customers' RLP that were obtained by using the three clustering methods. The MAE itself is calculated as follows:

$$MAE = \frac{\sum_{i=1}^{k} \sum_{j=1}^{nC_i} \frac{\sum_{t=1}^{96} \left| l_t^{TLP_i} - l_t^{RLP_{ij}} \right|}{96}}{\sum_{i=1}^{k} nC_i}$$

nC_i : The number of customers belonging to the ith cluster

$l_t^{TLP_i}$: The quarter-hour usage of TLP for the ith cluster

$l_t^{RLP_{ij}}$: The real quarter-hour usage of the jth customer belonging to the ith cluster

Hierarchical clustering uses 'similarity' as the comparison category to perform clustering. Thus, the technique is relatively less inefficient in minimizing the error distance, and its MAE is the largest compared to the other clustering methods. The k-means and fuzzy c-means clustering techniques use 'distance' as the comparison category to carry out clustering. With a smaller number of clusters, k-means clustering results in a similar MAE value as with fuzzy c-means clustering. With an increasing number of clusters, however, the method results in a somewhat smaller MAE value compared to fuzzy c-means clustering.

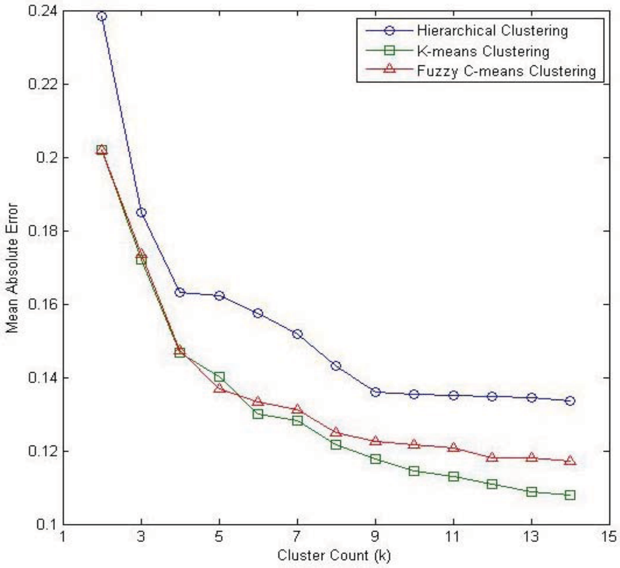

Figure 10. MAE in relation to cluster count

Based on the test results, two of the three clustering methods were found to offer benefits for this study. That is, with hierarchical clustering, researchers can ensure dozens of clusters and faster processing time despite its tendency to produce larger MAE values. With k-means algorithm, shorter processing time and smaller MAE values are guaranteed as long as approximately less than 10 clusters are used.

V. CONCLUSIONS

This study tested various research methods that are being used in generating typical load profile (TLP), which is one of the most fundamental elements in establishing and offering smart grid-based applications to energy consumers. To that end, 3,183 high-voltage (HV) customers of KEPCO were selected, and their AMR (automatic meter reading) data were used to perform clustering, with hierarchical, k-means, and fuzzy c-means (FCM) clustering adopted as specific methods to calculate the processing time required for clustering, and the MAE (mean absolute error) between TLP and RLP (real load profile).

According to experimental data analysis, the use of hierarchical clustering was found to be the optimal approach to ensure consistent processing time regardless of the cluster count. The k-means clustering technique, on the other hand, was judged to be the most efficient option if minimizing the MAE value between TLP and RLP is the overriding priority.

ACKNOWLEDGMENT

This work was supported by the research project (Development of integrated solution for continuous daily demand response based on demand estimation) of KEPCO Research Institute funded by the KEPCO.

REFERENCES

[1] J. W. Hughes, and D. W. Von Dollen, "Developing an integrated energy and communications systems architecture: the initial steps," in Proc. 2004 IEEE Power Systems Conference and Expositions, pp.1651-1654, 2004.

[2] Seon Gu Kim, Geong Il Hur, and Yeoung-Jin Chae, "Smart grid and its implications for electricity market design", Journal of Electrical Engineering and Technology, vol. 5, no. 1, pp. 1-7, March, 2010.

[3] SanJeev Kumar Aggarwal, Lalit Mohan Saini, and Ashwani Kumar, "Electricity price forecasting in ontario electricity market using wavelet transform in artificial neural network based model", International Journal of Control, Automation, and Systems, vol.6, no.5, pp.639-650, October 2008.

[4] Jain A. K. and Dubes R.C., Algorithms for Clustering Data, Englewood Cliffs, NJ: Prentice-Hall, 1988.

[5] Teemu Rasanen, Dimitrios Voukantsis, Harri Niska, Kostas Karatzas, and Mikko Kolehmainen, "Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data," Applied Energy, vol. 87, pp. 3538-3545, 2010.

[6] Iswan Prahastono, D. King, and C.S. Ozveren, "A review of electricity load profile classification methods," Universites Power Engineering Conference 2007, pp. 1187-1191, 2007.

[7] Bezdec, J.C., Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York, 1981.