Development of Representative Curves for Classified Demand Patterns of the Electricity Customer

In Hyeob Yu*, Jin Ki Lee*, Jong Min Ko* and Sun Ic Kim*

* Korea Electric Power Research Institute , Daejeon, Korea
(Tel : +82-42-865-5956; E-mail: ihyu, jklee, jmko, sikim@kepri.re.kr)

Abstract: Introducing the market into the electricity industry lets the multiple participants get into the new competition. These multiple participants of the market need new business strategies for providing value added services to customer. Therefore they need the accurate customer information about the electricity demand. Demand characteristic is the most important one for analyzing customer information. In this study load profile data, which can be collected through the Automatic Meter Reading System, are analyzed for getting demand patterns of customer. The load profile data include electricity demand in 15 minutes interval. An algorithm for clustering similar demand patterns is developed using the load profile data. As results of classification, customers are separated into several groups. And the representative curves for the groups are generated. The number of groups is automatically generated. And it depends on the threshold value for the distance to separate groups. The demand characteristics of the groups are discussed. Also, the compositions of demand contracts and standard industrial classification in each group are presented. It is expected that the classified curves will be used for tariff design, load forecasting, load management and so on. Also it will be a good infrastructure for making a value added service related to electricity.

Keywords: Demand Pattern, Load Profile, Customer Classification, Typical Load Profile, Automatic Meter Reading

1. INTRODUCTION

As the business environment of the electricity industry changes, the demand patterns of electricity customer become important for the market participants in the industry. Introducing the market into the electricity industry gives a new competition between the multiple participants. These multiple participants of the market need new business strategies for customer services because of the new competition. Therefore, customer information becomes crucial for making marketing strategies in the electricity industry. That is, establishment of the good strategies depends on what and how much they collect from electricity customer. And the analysis of customer information is essential for providing the differentiated services to customer. Demand characteristic is the most important one for analyzing customer information. Until now, demand characteristic has been studied using monthly billing data. But billing data is not enough for analyzing the demand characteristic. Recently Automatic Meter Reading (AMR) system has been introduced into Korea Electric Power Corporation (KEPCO). The number of customer included in AMR system would be about 100,000. Most of them would be the commercial and industrial customers who have contracts over 100 kW demand. In the future, the AMR system will be expanded to customers who have over 50 kW demand. The data collected from AMR include load profiles or interval data, which have the demand records for time intervals. Electronic power meters of KEPCO system have demand records for 15 minutes interval. Thus load profile data have the 96 intervals for demand history in a day, and the value in each interval represents electric demand for 15 minutes. In this study, the load profiles are classified into several groups such that similar patterns belong to the same group.

An algorithm for classifying the load profiles is developed for load analysis. Several procedures are needed for classifying load profiles. First, the interval data are averaged over a period of time duration according to day type. A metric is used for measuring the distance between groups. Next, the load profile data are normalized for the comparison purpose in the stage of preprocessing. Also, a feature vector, which is composed of maximum, minimum, average in demand, is generated for each profile. A combined use of supervised and unsupervised learning method[1] is used for classification. The number of groups is automatically generated. And it depends on the threshold value for the distance to separate groups. Thus, the load profiles in a same group have similar characteristics of electric demand.

As results of classification, representative curves for the same groups are generated. using the developed algorithm. The demand characteristics of the classified groups are discussed. Also, the cumulative demands of all groups are generated to show the contribution of the peak demand. The compositions of demand contract and standard industrial classification in each group are presented. Most of load research use the sampled data from customer[2], but the study uses all the load profile data from the AMR system.

2. DATA PREPROCESSING

The electric power meter has a memory for storing 15 minutes interval data, which are recorded from counting the pulses proportional to the electrical demand. Korea Electric Power Corporation (KEPCO) has built the AMR system since 2000. The automatic meter reading (AMR) system collects the data from the power meters every one hour through wireless communication networks. The system has 15 local servers in the branch offices of KEPCO. There exist about 100,000 customers who are included in the AMR system. Most of them are commercial and industrial customer. Each meter in the AMR system accumulates 96-point data for interval demands daily which is called the load profile data. The load profile data accumulated only in one day becomes huge in volume for all customers.

The data can be contaminated in the process of delivery from customer facilities to the server located in headquarter of KEPCO. Meters may have errors in storing data due to wrong operations. And there may exist some errors in data transfer between meters and the collecting center and in the processing for changing format of the collected data. Also there some errors in customer information that is usually updated daily because customer status is changing every day.

Thus preprocessing of the data is needed to select effective data for analysis. Several rules for preprocessing are given below.

If intervals with no records exceeds 20% of the entire analyzing period of time, the customer corresponding to the record will be discarded from the analysis database. That is, the customer will be deleted from the eligible list for the load analysis. If the outlier of the daily profile is over $\pm 3\sigma$ of the average vale of the daily profile, the data will be discarded.

Among eligible customer data, there exist missing parts of the profile intervals. If the continuous intervals less than 2 hours are missing, the missing part of the data will be estimated using linear interpolation. If the continuous intervals greater than 2 hours are missing, the missing part of the data will be estimated using the historical data. The historical data should be selected from the same type of day, month, and season before the data to be estimated.

3. CUSTOMER CLASSIFICATION

The load profile, demand information for customer, is composed of 96 intervals for one day. Each interval has the electricity demand value in the 15 minutes. To analyze demand patterns, it is necessary to classify the customer load profiles into several groups, which have similar demand profiles. An algorithm for classify the customer load profiles is presented.

To be prepared for deriving demand patterns, the load profile data are normalized between 0 and 1. Then daily load diagrams are derived for each customer.

Classification process uses some indices related to the diagrams. The indices related to the classification are defined as follows. First of all, a set of load diagrams is defined as $l = \{l_h \; h = 1,\dots,H\}$.

And load diagram of a specific group with M customers would be $L = \{l^{(m)} \; m = 1,\dots,M\}$.

If M customer would be classified into K groups, $L^{(k)}$, a sub set of L, includes $n^{(k)}$, where $k = 1,\dots, K$.

The representative curve of each classified group is defined for $\mathrm{r}^{(k)}$. The curve, $\mathrm{r}^{(k)}$, can be obtained from averaging the load diagrams in the $\mathrm{L}^{(k)}$.

Thus, a set of all the representative curves $\mathrm{R} = \{\mathrm{r}^{(k)} \; \mathrm{k} = 1,\dots,\mathrm{K}\}$ can be derived.

To separate load diagrams into groups, the metric standards for measuring the distance between load diagrams are defined as follows.

- Distance between two load diagrams

$$d(l^{(i)}, l^{(j)}) = \sqrt{\frac{1}{H}\sum_{h=1}^{H}(l_h^{(i)} - l_h^{(m)})^2}$$

- Distance between $L^{(k)}$ and a representative load diagram

$$d(r^{(k)}, L^{(k)}) = \sqrt{\frac{1}{n^{(k)}}\sum_{m=1}^{n^{(k)}} d^2(r^{(i)} - l^{(m)})^2}$$

- Distance between $L^{(k)}$ and a fixed load diagrams

$$d(l^{(i)}, c) = \sqrt{\frac{1}{H}\sum_{h=1}^{H}(l_h^{(i)} - c)^2}$$

- Distance between L and a customer load diagram, M

$$\hat{d}(L) = \sqrt{\frac{1}{2M}\sum_{m=1}^{M} d^2(l^{(m)} - L)}$$

And indices to derive demand characteristics from many load diagrams are needed. The indices are equivalent to a kind of feature vector in signal processing. Those indices are usually such as non-uniform coefficients, fill-up coefficients that are made up of maximum, minimum and mean values and modulation coefficient that indicates contribution of peak load and off-peak load. The developed algorithm for classification does not need the number of groups for initializing the program at first. Also the algorithm automatically derives the center of the classified groups. A weighting factor is used for the indices. It is defined for $\sigma_j^2 / \bar{\sigma}^2$, where σ_j is the standard deviation of the j[th] index, $\bar{\sigma}$ is the mean of the standard deviation of the indices. The effect of the weighting factor will be enlarged if the value of the standard deviation is big. Fig. 1 shows the schematic diagram for the classification algorithm. First, the normalized load profiles for all customers are prepared for classification. The feature vectors are derived for all customers according to the above definition. Minimum distance between groups is selected for the threshold value to separate groups. The distances from a customer load diagram to all the groups are measured. The minimum of the distances is determined for choosing the groups. If the minimum distance is great than the threshold value, the load diagram will be in a new group. If the minimum distance is less than the threshold value, the load diagram will be included the group which has the shortest distance from the customer load diagram.

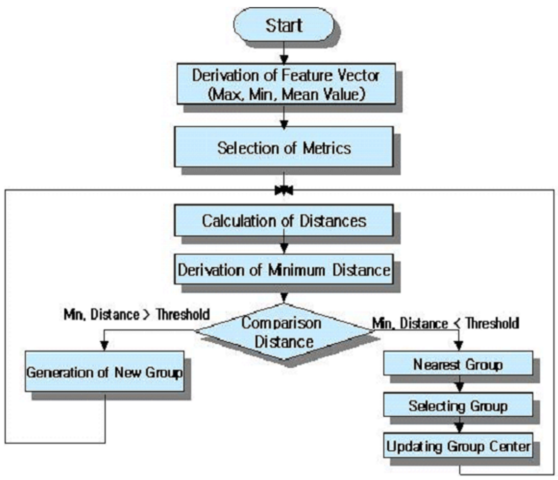

Fig. 1 Schematic diagram of classification algorithm

Then the center of the group should be updated since a new diagram comes into the group. Thus the number of groups and centers of the groups are determined. Repeated procedures are needed for updating the center of groups until no group is changed.

4. ANALYSIS OF CLASSIFIED GROUPS

The data used for this study come from the AMR system of KEPCO. The data include the billing, the demand and the customer information. AMR system has about 100,000 customers. And most of them are the commercial and industrial customers. All of the load data are used for classification without sampling.

4.1 Data for classification

4.1.1 Customer information

Data for the study is the regular information including customer identification and regular metering date and so on. Also the data has the billing information.

4.1.2 Demand information

Demand information comes from load profile data from AMR system. The data has demand value for 15 minutes interval.

4.1.3 Time period

Data available for this study is from July 2003 to June 2004. Also, the data is being accumulated once in 3 months.

4.1.4 Number of customers

The AMR system of KEPCO includes about 100,000 customers. All of the customer load profiles were used for analysis except the data failed in preprocess.

4.1.5 Database

Oracle Database 10g Standard Edition, which has automated storage management function, was used for the study.

4.1.6 Day-type

The load profiles were grouped into three day-types, such as weekdays, weekend, holiday.

4.2 Results of classification

Using the above data, the developed classification algorithm was tested for the weekday in July 2003. Fig. 2 shows the result of classification of the load profiles.

In Fig. 2, the abscissa indicates one-hour interval and the ordinate indicates demand magnitude, which is normalized to classify the groups. Eight demand patterns are shown in Fig. 2 as results of classification. Numbers of load diagrams for each demand pattern are shown in the legend box of the graph.

From the Fig. 2, the characteristics of the demand pattern are represented in the hourly demand behavior. The eight demand patterns are categorized into three different kinds of demand behavior.

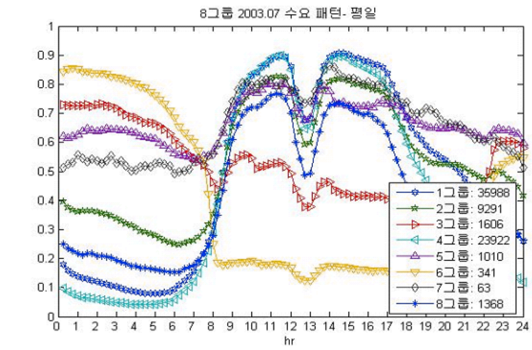

Fig. 2 Demand Patterns for 2003.07(Weekdays)

Group 1, 2, 4, 8 are indicated high demand at daytime. Group 3, 5 7 are indicated flat demand all day long. Group 6 shows high demand at night.

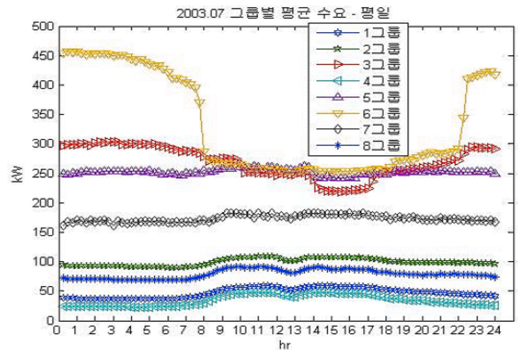

Fig. 3 Demands of Classified Groups 2003.07(Weekdays)

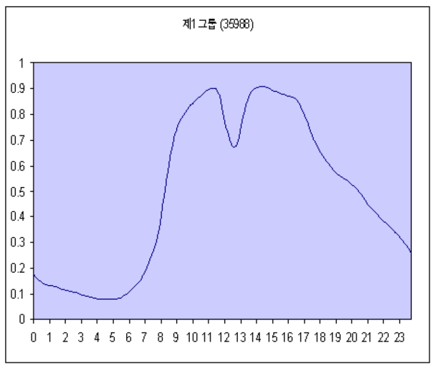

Fig. 4 Demand Pattern of Group 1 2003.07(Weekdays)

Fig. 3 shows average demand of each classified group. In the figure the ordinate of the graph indicates the demand value for each group in 15 minutes interval.

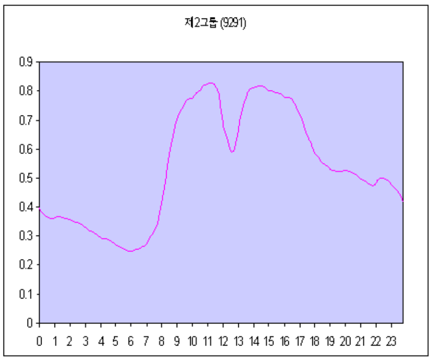

Fig. 5 Demand Pattern of Group 2 2003.07(Weekdays)

Demand pattern of Group 1 is shown in Fig. 4. The lower demand in the lunchtime for Group 1 is seen from the single demand graph in detail. And demand pattern of Group 1 is shown in Fig. 5. Next, information of customer for each group is analyzed.

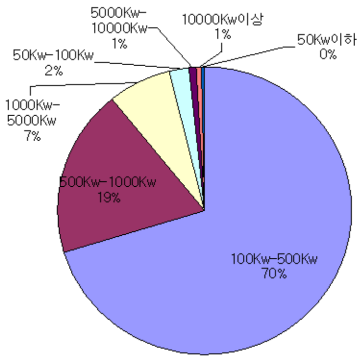

Fig. 6 Contractual Demand of Group 1

Fig. 6 shows that customers in Group 1 have different demand contracts with KEPCO. The figure indicates many different contractual demands have similar demand pattern. 70 % of customers in the group have demand contracts between 100kW and 500 kW. They are typical industrial customer.

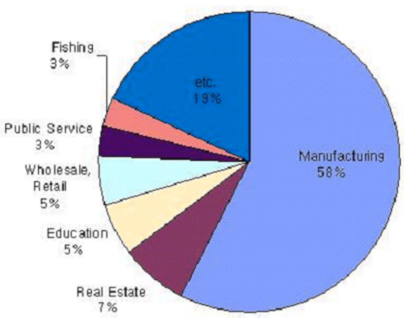

Fig. 7 Contractual Demand of Group 2

Fig. 7 shows the Standard Industrial Classification (SIC) of customers in Group 2. Many different business types are shown in Group 2. Thus SIC does not lead to demand pattern classification.

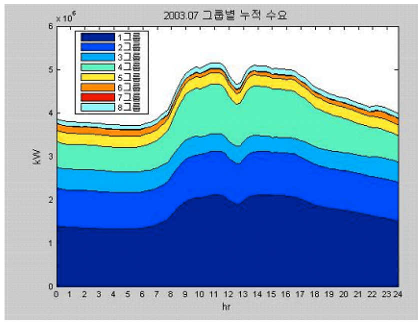

Fig. 8 Accumulated Demand of All Groups

Fig. 8 represents accumulated demands of all customers in each group in July 2003. Thus the area for each group indicates the actual energy usages in weekdays. Therefore the data can be used for analysis of contributions for peak demand in hourly interval.

KEPCO has several different categories for classify the electricity customer such as contractual demand, contractual business type, SIC, service type and so on. Those classifications have been used for billing process. Rate design shall use the data from actual demand patterns. But AMR systems have recently been installed in the fields, and the load profile data are available for industrial customer.

In the following, compositions of demand patterns in present categories are investigated using the classification algorithm. All customers who have contractual demands between 50kW and 100 kW are classified. Fig. 9 shows the result of the classification.

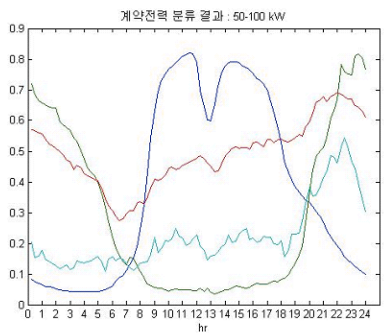

Fig. 9 Pattern Composition of Demands in 50-100 kW

It can be seen from Fig. 9 that four demand patterns are mixed in the group having the contractual demand between 50kW and 100kW. From the results, the classification of contractual demands identifies four different demand patterns in the category.

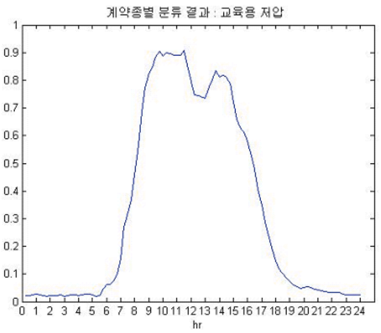

Fig. 10 Demand Pattern for Educational Contract

Fig. 10 and Fig. 11 show the demand patterns of customers who have contracts with KEPCO according to their business types. The classification algorithm has been applied to educational institutes. The number of the final classified group is only one as in Fig. 10. This means that educational institutes have a quite homogeneous demand pattern. In contrast, general use contract group has 5 different demand patterns.

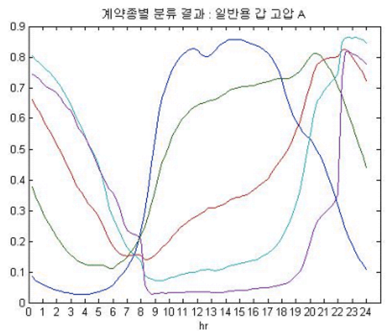

Fig. 11 Demand Pattern for General Use Contract

Therefore the group is made up of customers who have many different demand behaviors. Fig. 12 and Fig. 13 show the results of classification on the commercial service and an apartment complex. Commercial service loads include many different demand patterns.

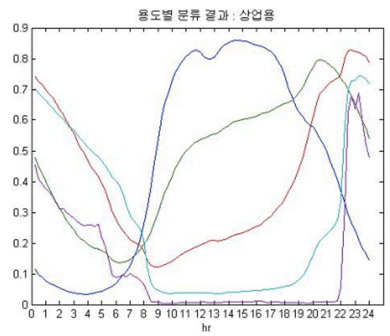

Fig. 12 Demand Pattern for Commercial Service

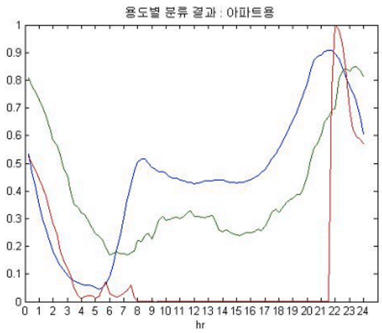

Fig. 13 Demand Pattern for Apartment Complex

But the apartment complex shows very similar demand patterns. Considering the data was taken from at July, it can be seen that home appliances are working before noon to avoid working at high temperature. When all family members come back from working, the demand goes up with fans, air-conditioners at night.

5. CONCLUSION

Traditionally, load research has been done with monthly billing information with statistical approaches. Nowadays, many studies on load profile data have been done[3]. In this work, demand patterns in the 15 minutes interval are classified with the developed algorithm. The data was taken from AMR system of KEPCO. The classified demand patterns will be an infrastructure for analyzing actual load behaviors on demand side. Also the demand patterns will be important data for the utility in the analysis of customer information. In the future, temperature sensitivity of the classified curves will be developed. It is expected that the classified curves will be used for tariff design, load forecasting, load management and so on. Also it will be a good infrastructure for making a value added service related to electricity.

REFERENCES

[1] Y.-H. Pao and D. J. Sobajic, "Combined use of unsupervised and supervised learning for dynamic security assessment," *IEEE Trans. Power Syst.,* vol. 7, pp. 878-884, May 1992.

[2] C. S. Chen, J. C. Hwang, Y.M. Tzeng, C.W.Huang and M.Y.Cho, " Determination of Customer Load Characteristics by Load Survey System at Taipower" *IEEE Trans. Power Delivery,* vol. 7, pp. 1430-1436, July 1996.

[3] Gerbec, D., Gasperic, S., Smon, I. and Gubina, F. "An approach to customers daily load profile determination" *Power Engineering Society Summer Meeting, 2002 IEEE,* Volume: 1, pp. 587-591 July, 2002