CLUSTER ANALYSIS FOR REGION ELECTRIC LOAD FORECASTING SYSTEM*

Hong Kyu Park[1], Young-il Kim[2], Jin Hyoung Park[1], Keun Ho Ryu[1†]
Database/Bioinformatics Laboratory, School of Electrical & Computer Engineering,
Chungbuk National University, Republic of Korea[1]
{hkpark1980, jhpark, khryu}@dblab.chungbuk.ac.kr[1]
Power Information Technology Group, Korea Electric Power Research Institute, Republic of Korea[2]
yikim @kepri.re.kr[2]

ABSTRACT: This paper is to cluster the AMR (Automatic Meter Reading) data. The load survey system has been applied to record the power consumption of sampling the contract assortment in KEPRI AMR. The effect of the contract assortment change to the customer power consumption is determined by executing the clustering on the load survey results. We can supply the power to customer according to usage to the analysis cluster. The Korea a class of the electricity supply type is less than other country. Because of the Korea electricity markets exists one electricity provider. Need to further divide of electricity supply type for more efficient supply. We are found pattern that is different from supplied type to customer. Out experiment use the Clementine which data mining tools.

KEY WORDS: Clustering, Load Pattern, K-means, Forecasting

1. INTRODUCTION

The other country advent of competitive electricity markets, the distribution side of the electricity industry has to face new challenges in providing satisfactory service to customers. The big challenges, there is the constant pressure for continuously decreasing the distribution service costs, which eventually reflects on the satisfaction of the supplied customers. Taking the customers' electrical behaviour into account is a key factor for setting up new tariff offers, leading to tariff structures more closely related to the actual costs of electricity provision in different contract assortment.

Therefore, the other country electricity provide have various tariff system. The customers supply the electricity satisfactory service.

But, the Korea electricity markets do not competition. One electricity provider very hard provides satisfactory service to customers. It is waste time and money. For don't waste time and money can be use the data mining.

The researchers' attention has been attracted by the load pattern issues for quite some time, seemingly less effort has been put in the analysis of available tools to produce a fair, sound classification system able to provide distinct and non overlapping customer classes.

[2] is investigate the potential of air conditioning load management by solving the temperature sensitivity of load demand for various customer classes. It also analyzes the load pattern. Other research is [3]. It compares the result obtained from the two approaches by means of two suitably defined adequacy indicators and discusses the potential applications of the surveyed approaches.

In this paper focus to divide the contract assortment and forecasting the pattern where can be import made classes.

We used the Clementine of data mining tools from SPSS. There are 2 clustering algorithm such as K-means [1] and Two step [4].

In our experiment use the K-means algorithm. It is fast and efficient more than Two step algorithm for sequential data.

2. DATA PREPARATION AND CLUSTERING MODELS

In this section, we describe how to make the datasets and clustering algorithms.

The load pattern associated with any customer contains the information of contract assortment, supply types and capacity which recoded every 15 minutes.

We consider a population of M customers, each customer being represented by a load pattern consisting of a group of H data $l = \{l_h, h = 1, \ldots, H\}$. We denote by $L = \{l^{(m)}, m=1, \ldots, M\}$ the set of load patterns associated to the M customers.

2.1 Data Preparation

We use 231 customers AMR data for March 2007 in KEPRI AMR. It has 31 days and measured an interval 15 minutes from 100 to 2345. It is change to the values into the cluster algorithm in Clementine.

First, we group to the all date by ascending order and also group to the Customer ID.

† Corresponding Author

Second, calculate the average for values that recorded every 15 minutes.

We use pre-processing data into the clustering algorithm in Clementine.

2.2 K-means Clustering Models

The K-means clustering technique is simple. The K-means first choose K initial centroids, where K is a user-specified parameter, namely, the number of clusters desired. Each point is then assigned to the closest centroid, and each collection of points assigned to a centroid is a cluster. The centroid of each cluster is then updated based on the points assigned to the cluster. The K-means repeat the assignment and update steps until no point changes clusters, or equivalently, until the centroids remain the same.

K-means is formally described by Algorithm (). The operation of K-means is illustrated in Figure 1, which shows how, starting from three centroids, the final clusters are found in four assignment-update steps.

The centroids are indicated by the "+" symbol; all points belonging to the same cluster have the same marker shape.

Algorithm 1. Basic K-means algorithm.

1: Select K points as initial centroids.
2: **repeat**
3: Form K clusters by assigning each point to its closest centroid
4: **until** Centroids do not change.

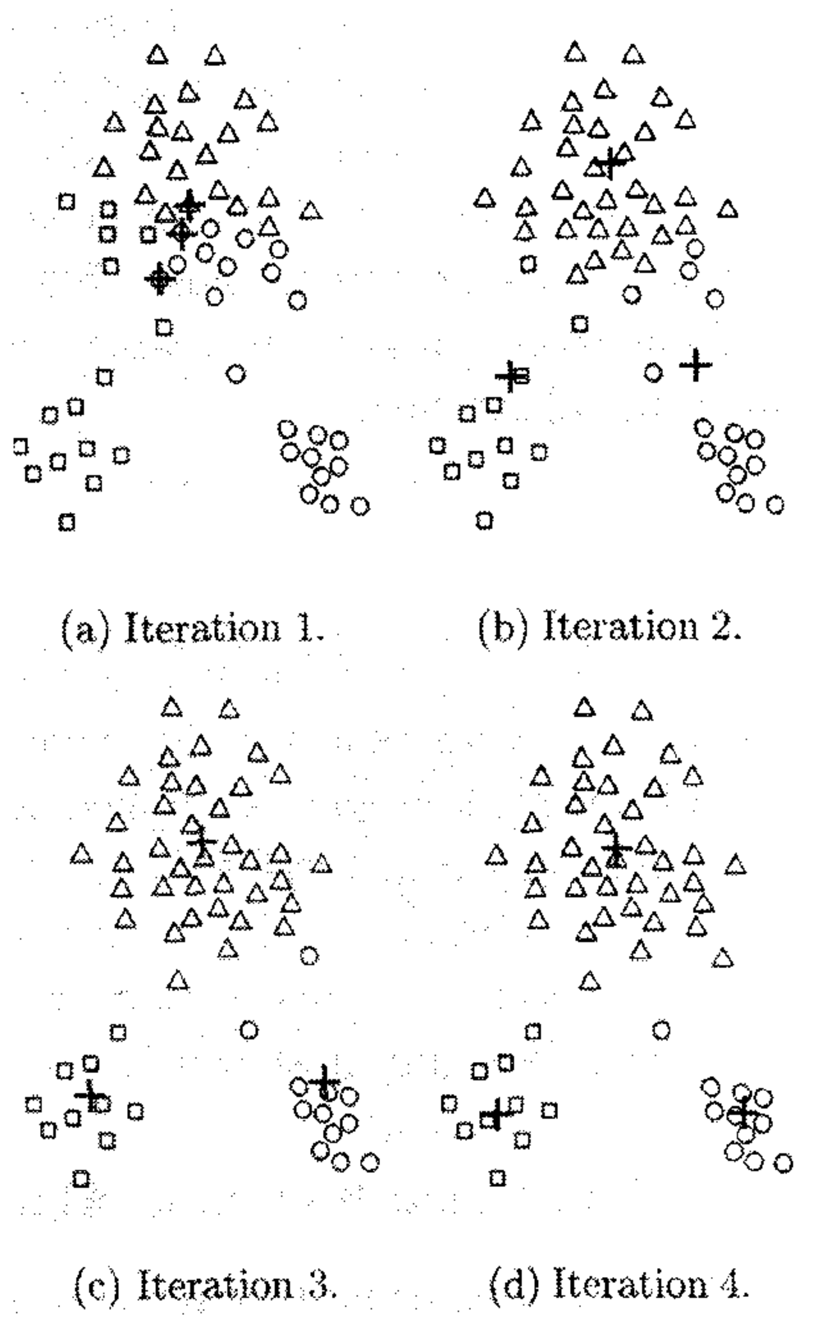

Figure 1 Using the K-means algorithm to find three clusters in sample data

First strategy, shown in Figure 1, points are assigned to the initial centroids, which are all in the larger group of points. For this example uses the mean as the centroid. After points are assigned to a centroid, the centroid is then updated. Again, the figure for each step shows the centroid at the beginning of the step and the assignment of points to those centroids. In the second strategy, points are assigned to the updated centroids, and the centroids are updated again. In strategy 2, 3 and 4, which are shown Figures 1(b), (c), and (d) respectively, two of the centroids move to the two small groups of points at the bottom of the figures. When the K-means algorithm terminates in Figure 1(d), because no more changes occur, the centroids have identified the natural groupings of points.

Table 1. Table of notation 1

Symbol	Description
X	An object.
C_i	The i^{th} cluster.
c_i	The centroid of cluster C_i.
c	The centroid of all points.
m_i	The number of objects in the i^{th} cluster.
m	The number of objects in the data set.
K	The number of clusters.

The K-means calculate the error of each data point using the sum of the squared error (SSE)

Using the notation in Table 1, the SSE is formally defined as follows:

$$SSE = \sum_{i=1}^{k} \sum_{x \in C_i} dist(c_i, x)^2 \tag{1}$$

where *dist* is the standard Euclidean distance between two objects in Euclidean space.

Using the notation in Table 1, the centroid (mean) of the ith cluster is defined by Equation (2).

$$c_i = \frac{1}{m_i} \sum_{x \in C_i} X \tag{2}$$

3. LOAD PATTERN ANALYZE

After the clustering using the clementine we obtain the text value adding the cluster labels. We copy the text data and attach into the excel. Using the excel, we drawing the graph and separate the each cluster.

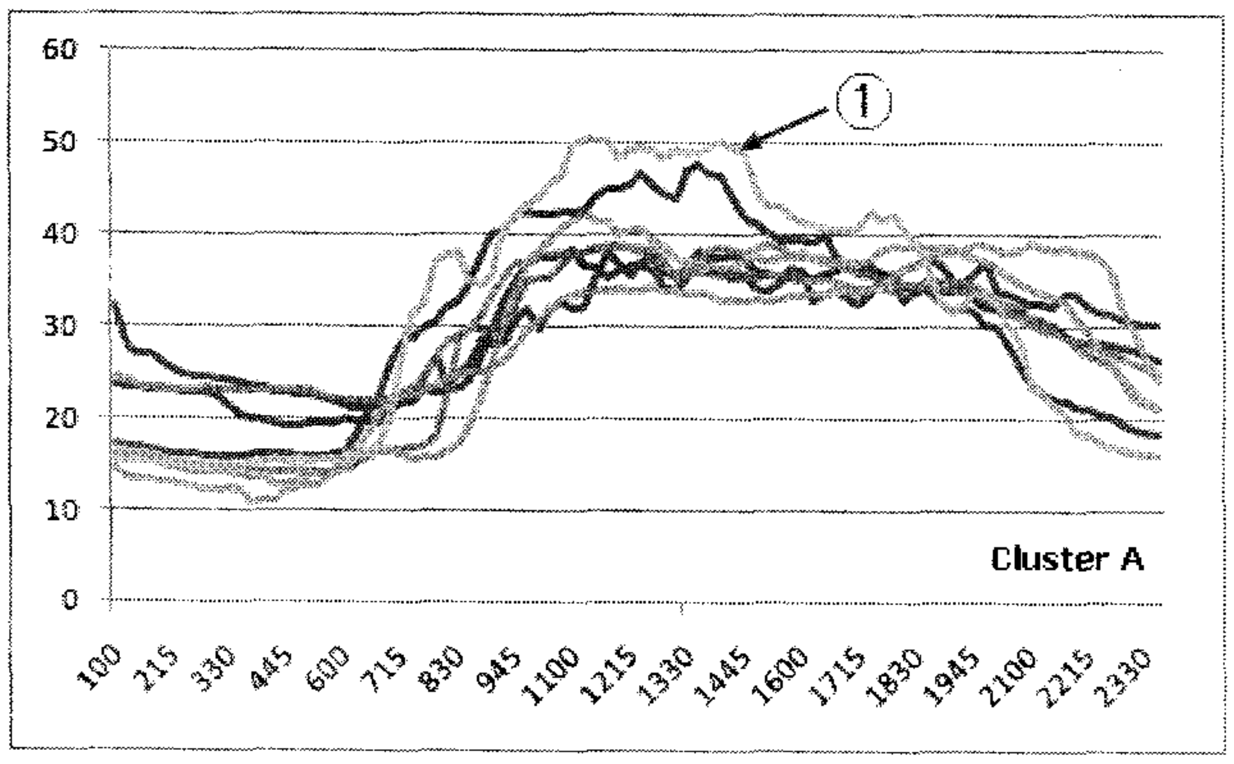

Figure 2. The load pattern of the Cluster A

In Figure 2, the direct the red arrow pattern's contract assortment is 226 which mean general usage electricity (gap) other pattern contract assortment is 221 which mean general usage electricity (eul). A contract assortment 226 pattern is very similar a contract assortment 221, so a contract assortment 226 can be supply like to a contract assortment 221.

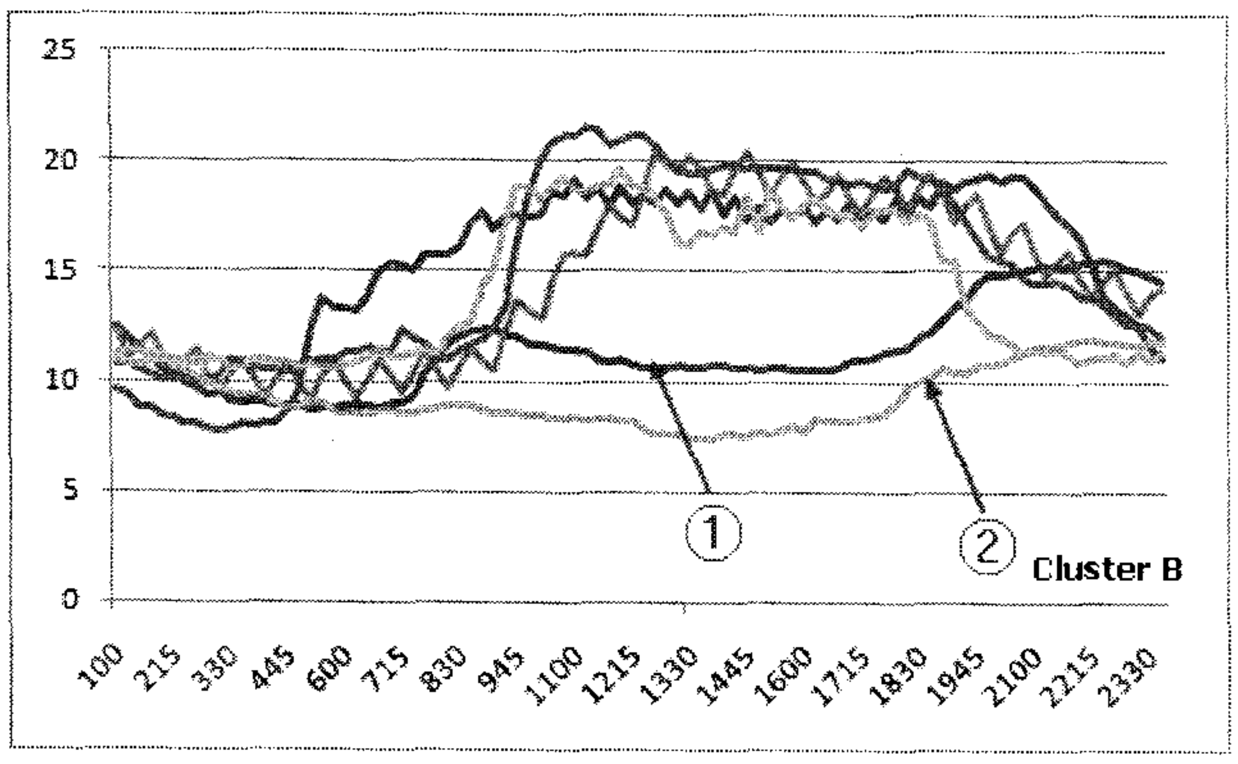

Figure 3. The load pattern of the Cluster B

The Figure 3 cluster shows a contract assortment 221. But No.1 graph is a contract assortment 110. In this case, No.2 graph very similar a contract assortment 110. So, the electricity provide can be supply like to 110 instead of No.2 customer.

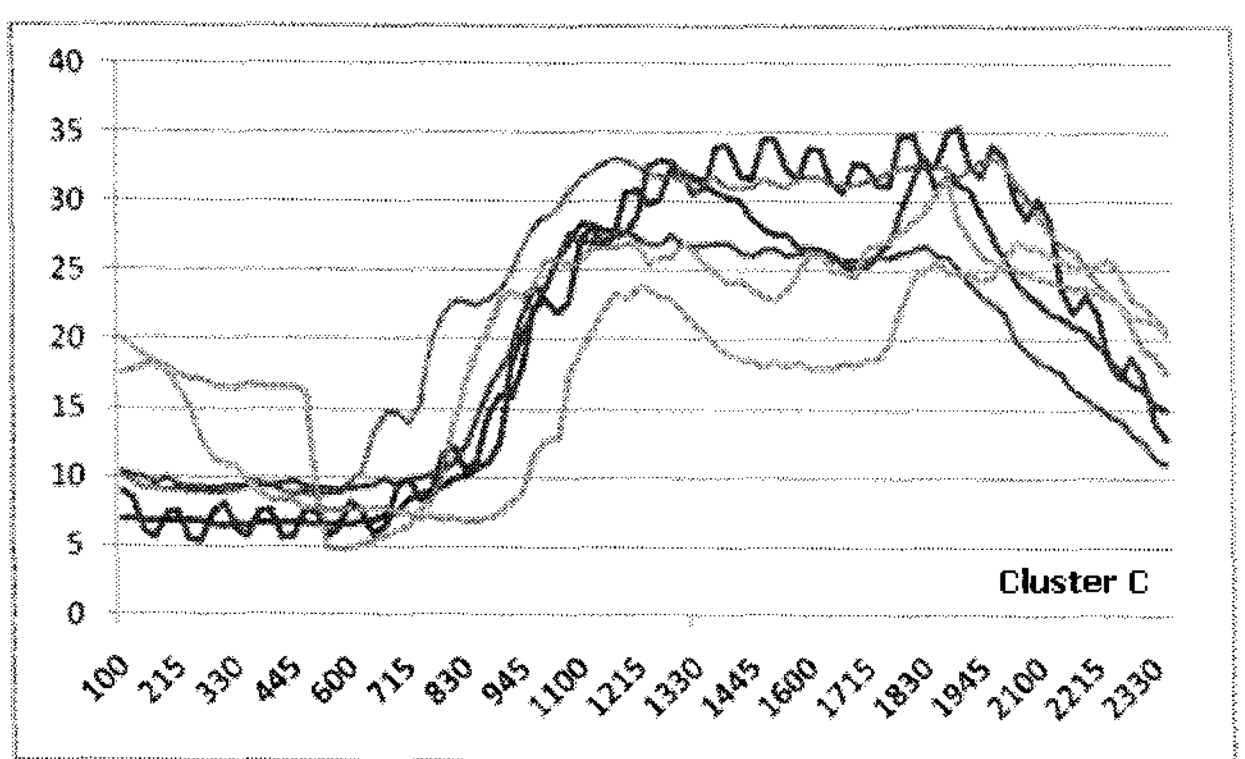

Figure 4. The load pattern of the Cluster C

Also, we can forecast input new AMR data. Figure 4 show a contract assortment 221. When the input new AMR data if the new AMR data pattern like to cluster C, the electricity provider can be supply a contract assortment 221 type to new customer.

As shown the Figure 2, 3 and 5 are a contract assortment 221 but, they are not equal the pattern. So, the electricity provides have to divide a contract assortment more than now then expecting efficient electricity supply.

4. CONCLUSION

In this paper, we found the customers supply the electricity dissatisfactory service load pattern and forecasting where new input AMR data into the contract assortment. The test customers extract to record in load survey system in KEPRI.

We found the dissatisfactory service to customers and contract assortment is different but load pattern is similar.

More efficient the electricity satisfactory service to customer, need to more contract assortment type. And about new customer can be supply the electricity service do not waste time and money.

REFERENCES FROM OTHER LITERATURE:

[1] Andrew R. Conn, Luís N. Vicente, (1999) *Two-Step Algorithms for Nonlinear Optimization with Structured Applications,* Chandu Visweswariah

[2] G. Chicco, R. Napoli, F. Piglione, P. Postolache (2004), M. Scutariu, C. Toader, *Load Pattern-Based Classification, IEEE TRANSACTIONS ON POWER SYSTEMS,* VOL. 16, NO. 2

[3] J. B. MacQueen (1967): *Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability,* Berkeley, University of California Press, 1:281-297.

[4] J. C. Hwang (2001), *Assessment of Air Condition Load Management by Load Survey in Taipower,* IEEE TRANSACTIONS ON POWER SYSTEMS, VOL. 16, NO. 4

ACKNOWLEDGEMENTS

*This research was supported by development of AMR system interfacing model on internet GIS environment project of the Korea Electric Power Research Institute (KEPRI)