전력 소비 데이터로부터의 출현 부분패턴 추출

박현우, 박명호, 류근호*
데이터베이스/바이오인포메틱스 연구실, 충북대학교
e-mail : {hwpark, bluemhp, khryu}@dblab.chungbuk.ac.kr

Emerging Subspace Discovery from Daily Load Consumption Data

Hyun Woo Park, Minghao Piao, Keun Ho Ryu
DB/Bioinformatics Lab., Chungbuk National University

Abstract

Customers of different electricity consumer types have different daily load shapes in the manner of different characteristics. Therefore, maximally capture load shape variability are desirable in load flow analysis. And most of time, such load shape variability can be found in the particular subspace of load diagrams. Therefore, in this paper, we are using subspace projection method to capture the emerging subspaces of load diagrams which maximize the difference between particular load shapes in different group of customers. As the result, subspace projection method can be used in load profiling and the performance is good as traditional approaches.

1. Introduction

The knowledge of how and when consumers use electricity is essential to the competitive retail companies. The knowledge can be found by data mining application to historical data of the consumers collected in load research projects. Clustering and load profiling [1, 2, 3, 4, 5, 6, 7], classification [8, 9, 10] and forecasting [11, 12, 13] has been main research topics during last years. The goal of load profiling is to partition the initial data in to a set of classes according to the load shape of the representative load diagrams of each consumer. This is made by assigning the consumers into same class with the most similar behavior, and consumers with dissimilar behavior into different classes [6, 12, 14, 15, 16].

Clustering is used to generate groups of data from a dataset with the intention of representing the behavior of a system as accurately as possible [8, 10]. Clustering algorithms also can be used to obtain a better support management decisions or achieve the segmentation and demand patterns for electrical customers on the basis of database measurements [9, 11]. Differences between an individual load profile and that of others within the same group can be used to suggest energy usage behavior changes to reduce overall electricity usage or to improve electrical efficiencies, possibly by shifting the usage time of particular appliances. In addition, for identifying the differences between load profiles, it needs to discover dimensions which values in these dimensions shows big differing shapes. This work can be done by using subspace projection based clustering algorithms.

In remains of paper, we introduce *Proclus* which is subspace projection method that can be used to find such dimensions most related to definition of load profiles in chapter 2, and give essential preliminaries in chapter 3 and describe our experimental results in chapter 4. In chapter 5, we make a summary and discussion of our study and give framework of our future work.

2. Subspace Projection Method

Subspace or projection method is an extension of traditional clustering algorithms that aims to find clusters in different subspaces of given fixed number of dimensions within a dataset. Subspace clustering algorithms may report several clusters for the same object in different subspace projections, while projected clustering algorithms are restricted to disjoint sets of objects in different subspace.

Proclus [17] is a clustering-oriented approach which needs to define properties of the entire set of clusters, like the number of clusters, average dimensionality or more statistically oriented properties. As they do not rely on counting or density, they are more flexible in handling different types of data distributions. *Proclus* partitions the data into *k* clusters with average dimensionality l, extending K-Medoids approach which called CLARANS [18]. The general approach is to find the best set of medoids by a hill climbing process, but generalized to deal with projected clustering. The algorithm proceeds in three phases: an initialization phase, an iterative phase, and a cluster refinement phase. The purpose of the initialization phase is to reduce the set of points and trying to select representative points from each cluster in this set. The second phase represents the hill climbing process that in order to find a good set of medoids. Also, it computes a set of dimensions corresponding to each medoid so that the points assigned to the medoid best form a cluster in the subspace determined by those dimensions. The assignment of points to medoids is based on Manhattan segmental distances relative to these sets of dimensions. Finally, cluster refinement phase, using one pass over the data in order to improve the quality of the clustering.

* Corresponding Author

3. Preliminaries

Emerging patterns [19] are special frequent patterns whose frequencies change significantly from one dataset to another. Each emerging pattern has big difference between its supports in the opposing classes and represents strong contrast knowledge. So, it can sharply differentiate the class relationship of input instances containing the emerging patterns. By using the skeleton of emerging patterns, emerging subspace is defined as:

Dimensions which are most related to definition of load profiles are called emerging subspaces. Emerging subspace represents the most differing shapes of load diagrams.

In figure 1, given two load diagrams have differing shape in emerging subspace where the subspace is most related to the definition of load profiles. In addition, these two load profiles are discriminated by the values in such particular dimensions. It indicates that definition of load profiles can be done by defining such emerging subspaces.

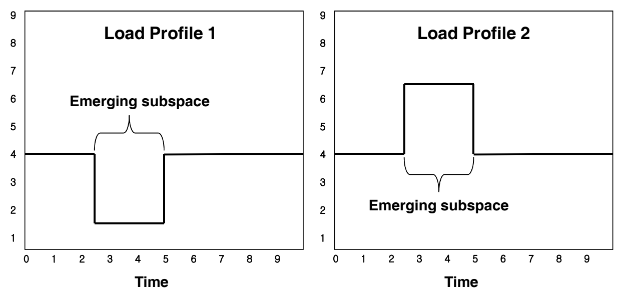

Figure 1. Example of load profiles which shows big difference in emerging subspace

4. Experimental Results

The measurements of individual commercial use load diagrams were performed by using Automatic Measurement Reader (AMR). The diagrams are measured at intervals of 15 min, therefore, resulting in 96 points in a day curve. In our study, we have aggregated the diagram from interval of 15 min into interval of hour for more efficient analysis as shown in formula 1, where for each consumer *c*, let $V^{(c)}$ denotes the total daily power usage of *c* in one day for 24 hours. And each instance is belonging to one of the 6 different contract types.

$$V^{(c)} = \{V_0^C, \ldots, V_h^C, \ldots, V_H^C\}, \text{c = customer, } 0 < h < 24, H = 24 \qquad (1)$$

Table 1 shows the parameter setting for *Proclus*. Since 6 contract types are used as class label information, number of clusters is given as 6, and we consider the number of dimensions from 1 to 24 for finding emerging subspaces which are most related definition of clusters. For performance evaluation, *Accuracy*, *Coverage*, *1.0-Entropy* and *F1*-value are used.

Accuracy: Accuracy of clustering is the degree of closeness of measurements of assigned cluster labels to that instances' actual cluster label.

Entropy and Coverage: Entropy accounts for purity of the clustering (values closer to zero indicate more complete clustering), while coverage measures the percentage of objects in any subspace cluster.

F1-value: The F1-value represents a harmonic mean between recall and precision. It is commonly used in evaluation of classifiers and recently also for subspace or projected clustering as well. In *OpenSubspace*, the F1-value of the whole clustering is simply the average of all F1-values.

Table 2 shows the clustering results and figure 2 shows the discovered emerging subspaces and load shapes in these subspaces. Obviously, it indicates that according to the clustering results it is possible to define load profiles for each cluster. Figure 3 shows the defined load profile diagram for each cluster.

<Table 1> Parameter Setting for *Proclus* in OpenSubspace

Method	Parameter	Fr	Offset (Op)	Steps	To
Proclus	Average dimensions	1	1 (+)	24	24
	No. of clusters	6	0 (+)	1	6
	Iteration : 10				
Total number of experiments: 240					

<Table 2> Evaluation Measurements for *Proclus*

Accuracy	0.82
Coverage	0.86
1.0-Entropy	0.73
F1	0.77

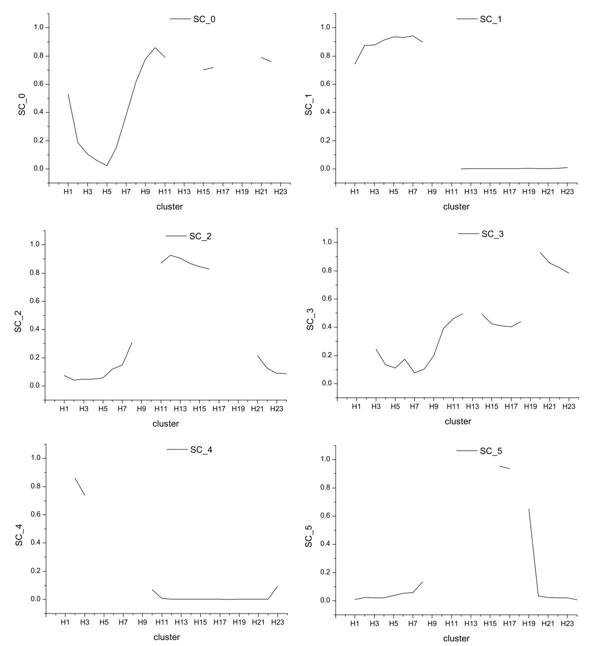

Figure 2. Emerging subspaces related to definition of clusters, and load shapes in emerging subspaces

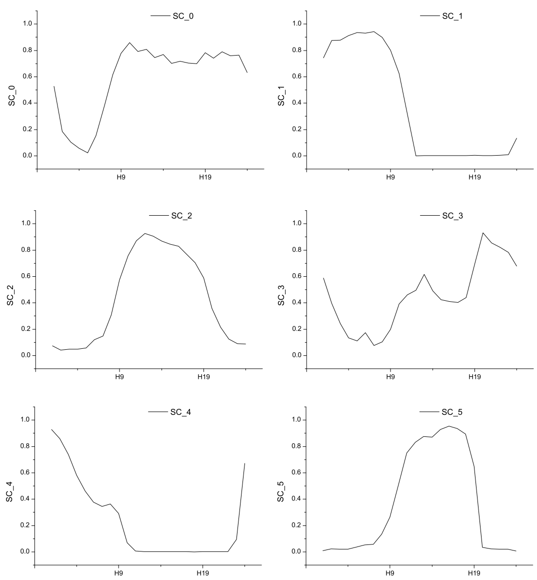

Figure 3. Defined load profile diagram for each cluster

5. Conclusion

Data mining application to electricity market is for better decision making and it is widely studied. Load profiling is one of the clustering applications to electricity market and traditional clustering algorithms like K-means and SOM are widely used. However, considering all dimensions is time consuming and it is possible to only consider subspaces to define clusters. Furthermore, load shapes represent big geometric differences in particular subspaces. Therefore, in this paper we used subspace projection method to find such subspaces and discovered their load trends in these subspaces. The result shows that it is possible to define load profiles for each group according to the clustering results.

6. Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2012-0000478)

References

[1] Z. Zakaria, "Cluster validity analysis for electricity load profiling," IEEE International Conference on Power and Energy (PECon), pp. 35-38, 2010.

[2] Y. I. Kim, J. M. Ko, J. J. Song, and H. Choi, "Repeated Clustering to Improve the Discrimination of Typical Daily Load Profile," *Journal of Electrical Engineering & Technology*, vol.7 no.3, 2012, pp. 281-287.

[3] F. Martínez-Álvarez, A. Troncoso, J.C. Riquelme, and J.M. Riquelme, "Partitioning-clustering techniques applied to the electricity price time series," The 8th international conference on Intelligent data engineering and automated learning, pp. 990-999, 2007.

[4] A. Gabaldón, A. Guillamón, M.C. Ruiz, S. Valero, C. Álvarez, M. Ortiz, and C. Senabre, "Development of a methodology for clustering electricity-price series to improve customer response initiatives," IET Generation, Transmission & Distribution, vol. 4, issue 4, pp. 706-715, 2010.

[5] S. V. Verdu, M. O. Garcia, C. Senabre, A. G. Marin and F. J. Garcia Franco, "Classification, filtering and identification of electrical customer load patterns through the use of SOM maps," *IEEE Transactions on Power Systems*, vol. 21, no. 4, 2006, pp. 1672-1682.

[6] G. Chicco, R. Napoli and F. Piglione, "Comparisons among Clustering Techniques for Electricity Customer Classification," *IEEE Transactions on Power Systems*, vol. 21, no. 2, 2006, pp. 933-940.

[7] J. H. Shin, B. J. Yi, Y. I Kim, H. G. Lee, and K. H. Ryu, "Spatiotemporal Load-Analysis Model for Electric Power Distribution Facilities Using Consumer Meter-Reading Data," *IEEE Transactions on Power Delivery*, vol. 26, no. 2, 2011, pp. 736-743.

[8] M. H. Piao, H. G. Lee, J. H. Park, and K. H. Ryu, "Application of Classification Methods for Forecasting Mid-Term Power Load Patterns," Communications in Computer and Information Science, vol. 15, Part 2, pp. 47-54, 2008.

[9] M. Ruska, S. Repo, and P. Jarventausta, "Customer Classification and Load Profiling Method for Distribution Systems," *IEEE Transactions on Power Delivery*, vol. 26, issue. 3, 2011, pp. 1755-1763.

[10] D. L. Huang, H. Zareipour, W. D. Rosehart, and N. Amjady, "Data Mining for Electricity Price Classification and the Application to Demand-Side Management," *IEEE Transactions on Smart Grid*, vol. 3, no. 2, 2012, pp. 808-817.

[11] S. Ye, G. Zhu, and Z. Xiao, "Long Term Load Forecasting and Recommendations for China Based on Support Vector Regression," Energy and Power Engineering, vol. 4, no. 5, 2012, pp. 380-385.

[12] R. Barzamini, F. Hajati, S. Gheisari, and M.B. Motamadinejad, "Short Term Load Forecasting using Multi-layer Perception and Fuzzy Inference Systems for Islamic Countries," *Journal of Applied Sciences*, vol. 12, issue. 1, 2012, pp. 40-47.

[13] S. Fan, and R. J. Hyndman, "Short-Term Load Forecasting Based on a Semi-Parametric Additive Model," *IEEE Transactions on Power Systems*, vol. 27, issue. 1, 2012, pp. 134-141.

[14] V. Figueiredo, F. Rodrigues, Z. Vale, and J. B. Gouveia, "An electric energy consumer characterization framework based on data mining techniques," *IEEE Transactions on Power Systems*, vol. 20, issue 2, 2005, pp. 596-602.

[15] I. Dent, U. Aickelin, and T. Rodden, "The Application of a Data Mining Framework to Energy Usage Profiling in Domestic Residences using UK data," Research Student Conference on Buildings Do Not Use Energy, People Do?, Bath, UK, 2011.

[16]G. Q Zhang, J. Lu, X. P Feng, and W. C. Yang, "A New Index and Classification Approach for Load Pattern Analysis of Large Electricity Customers," *IEEE Transactions on Power Systems*, vol. 27, issue. 1, 2012, pp. 153-160.

[17] C. Aggarwal, J. Wolf, P. Yu, C. Procopiuc, and J. Park, "Fast algorithms for projected clustering," ACM SIGMOD international conference on Management of data, 1999, pp. 61-72.

[18] R. Ng, J. Han, "Efficient and Effective Clustering Methods for Spatial Data Mining," *The 20th International Conference on Very Large Data Bases*, 1994, pp. 144-155.

[19] K. Ramamohanarao, J. Bailey, Hongjian Fan, "Efficient Mining of Contrast Patterns and Their Applications to Classification", Proceedings of the 2005 3rd International Conference on Intelligent Sensing and Information Processing, pp. 39-47, 2005.