APPLICATION OF SELF-ORGANIZING MAPS FOR CLASSIFICATION AND FILTERING OF ELECTRICAL CUSTOMER LOAD PATTERNS

Sergio Valero, Mario Ortiz, Carolina Senabre
Universidad Miguel Hernández
Avda. de la Universidad s/n, 03202. Elche, Spain
Email: svalero@umh.es, mortiz@umh.es

Carlos Álvarez, Fco. José García Franco, Nuria Encinas
Instituto de Ingeniería Energética, Valencia, Spain

Antonio Gabaldón, Juan A. Fuentes
Dpt of Electrical Engineering, Universidad Politécnica de Cartagena, Spain

Ignacio J. Ramírez-Rosado, Luis A. Fernández-Jiménez
Dpt of Electrical Engineering, Universidad de La Rioja, Spain

ABSTRACT

The objective of this paper is to show the capability of the Self-Organizing Maps (SOMs) to organize, to filter, to classify and to extract patterns from distributor, commercializer, aggregator or customer electrical demand databases -objective known as data mining-. This approach basically uses -to reach the above mentioned objectives- the historic load demand curves of each user. In our case, and for simplicity, we will study two typical medium users: an industry and a university located both in Spain. The results clearly show the suitability of SOM approach to improve data management and to find easily coherent clusters between electrical users.

KEY WORDS

Demand Management, Self-organizing maps, electricity markets, electrical customer segmentation, load patterns.

1. Introduction

The deregulation process began in developed countries a decade ago stimulated by political and technological reasons. Unfortunately the experience has not been as much successful as it was planed, due to a lot of problems appeared from 2000 up to now (California Energy Crisis in 2000, Blackouts in Europe, United States and Canada in 2003). Due to these experiences, regulators and system operators believe more and more that additional electricity resources -Distributed Energy Resources- should be procured using an integrated process that takes into account not only supply resources -Distributed Generation- but also some demand policies: for example efficiency gains in demand -in long term horizon- or price responsiveness -in short term horizon-. This supposes a new scenario where demand and supply compete on an equal footing in energy markets. For example, California Energy Commission will finance new energy efficiency programs to achieve a forecasted demand reduction of 6000 GWh in 2008 [1]. The effective contribution to these energy efficiency programs and the necessity of offering energy choices to consumers need a detailed knowledge of customer segments and the characterization of these clusters -from the point of view of energy uses-.
Besides, this new regulated framework of electrical power systems has promoted the necessity of new customer -and system- measurement, monitoring and control activities. This fact has increased the amount of data stored by supply-side actors. The enormous quantity of available data presents a problem for utilities but also a non negligible opportunity for distribution research. This high dimensional data set can not be easily modelled and advanced tools for synthesizing structures from such information are needed.
This is the main objective of the so called Data Mining techniques (DM) or more precisely Knowledge Discovery in Databases (KDD) [2], [3], and it is also the research purpose of this work applied to customer characterization and segmentation.
The paper is organized in five sections. In section 2 some principles of SOM networks are reviewed. Section 3 presents the characteristics of the two customer databases used in this paper, the criteria used for load data conditioning and the definition of time indices. Subsequently various SOM maps are obtained for the University customer collected data in order to evaluate the potential of Self-Organizing Maps in anomalous data filtering and in the search of different patterns in the customer behavior. Finally, we present a segmentation analysis for our two customers before and after an

anomalous data filtering process was performed -the details are described in section 5-. Finally some conclusions and future works are stated.

2. Self-Organizing Maps

The management of databases and the process that aims at extracting synthesized information from large amounts of data can be performed by a lot of techniques, such as relational statistical calculus -a traditional approach- or for example automatic learning -fuzzy inference or artificial neural networks-. This last approach has been selected for this application, and specifically Self-Organizing Maps (SOMs). The methodology was introduced by Teuvo Kohonen two decades ago [4]. These networks are a kind of unsupervised Artificial Neuronal Network (ANN) that performs a transform from the original input space -n dimensional data vector- to an output space -two dimensional in this case-. The advantage of SOM is that the relation between the original vectors is in some extend preserved in the two dimensional space, providing -through some analysis of these maps and the evaluation of some indices- a visual format that a human operator, with some expertise, allows him to “easily” discover clusters, relations and structures in these databases.

3. Data Processing and Conditioning

A. Generalities

The process of extraction of information from a database needs a previous treatment of the data to give them a uniform format that allows working with them. The daily load profiles present peaks and different values for each user. Therefore, it is necessary to normalize the data in order to have a common format for all the users and to allow their introduction in the neural network or the correspondent method of study avoiding the levels of demand in the first approximation.

Previous studies allowed the authors to achieve the conclusion that the most suitable format for data was fitting with load curves that presented 24 daily values of consumption [5], since we need to extract the more relevant characteristics of the user -with the minor cost of time and work- while considering the availability of the sources -measures usually taken by electrical companies-.

Finally, the comparison among customers can be realized through diverse types of indices and factors obtained from the above mentioned “standard” profiles. Two alternative approaches are used in the bibliography: Time Domain Indices and Frequency Domain Indices -defined using the Discrete Fourier Transform, DFT-, which allow to obtain additional information about consumption patterns [6], [7], [8].

B. SOM’s Application to Data Mining

As soon as the data have the same format, several tools can be applied to allow the filtering and detection of punctual values or undesired series. In this respect and as it appears in the present paper, the SOM seems a powerful tool able to find the possible presence of uncertainty in data –noise, missing information- that could exist in the database and profiles corresponding to days of abnormal consumption -holidays, days with reliability problems…- Even more, these networks permit to find standards of anomalous behavior owed to punctual changes in the habit of consumption –seasonal loads- and allow the definition of representative patterns of the customer across the average of the classified like “typical profiles”.

C. Case Study

The load daily profiles that compose the set of measurements for training and evaluation of SOM maps correspond to two typical medium users: an industry and an university located both in Spain. Data used in the training of the neural network correspond with weekly load curves -from Monday to Friday- of both customers measured in 2003. For simplicity Saturdays and Sundays profiles were not considered in this phase of study.

Table I shows the mask associated to daily load profiles -a number of mask for all the load profiles for a given customer-, used in the case of the classification of customers reported in section 5, which allows to identify the data assigned to a cell with the corresponding customer. The table also shows the number of load curves considered for each customer -including anomalous days later eliminated- that determines in certain way the size of the network.

TABLE I
CUSTOMERS SPECTRUM FOR SOM TRAINING

Customer	Label	N° of initial input data vectors	Customer activity
User 1	1	147	Medium Industry
User 2	2	195	University

An alternative labelling is used for data filtering purposes. This filtering is applied separately to both customers. By means of this labelling a number is assigned to each profile following the next criterion: the last two digits indicate the day of the month and the initial remaining ones the corresponding month (i.e., mm/dd). Thus, label maps that are obtained -see figure 2- allow to identify weekly load data assigned to each cell.

As previously stated, the use of information in Time Domain was chosen. Especially, the curves interpolated to 24 values of daily consumption that were used were obtained from load curves recorded every 15 minutes. The reason was simply the good results obtained in previous works accomplished by the authors [8].

4. Application of SOM for anomalous data filtering

In the process of result optimization, different possible configurations were proved for the parameters of the network. A map size of 16x16 cells and a number of 2000 and 1000 steps for primary and secondary training respectively was finally applied. Some results and characteristics observed for the User 2 -University- after a randomly feeding of the network are discussed.

The SOM training results obtained with weekly load profiles -from Monday to Friday- for User 2 are shown in the color map presented in figure 1.

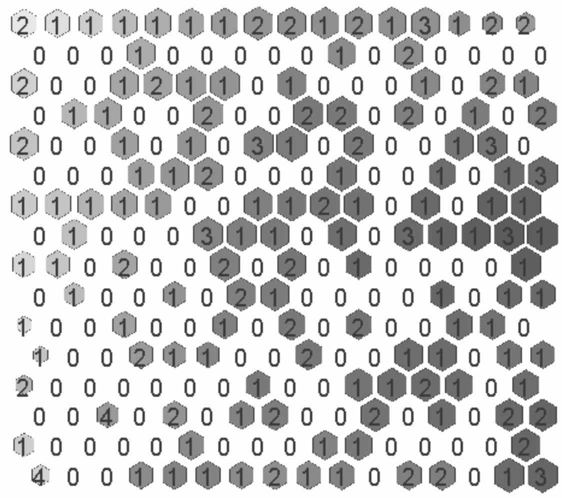

Figure 1. Color Code Map. University

In this one it is observed the number of profiles that were presented to the SOM and were assigned to the same cell marked with a digit. Likewise, the relative size of the above mentioned cells indicates the level of similarity of the profiles in relation to the rest of the set of data employed in the training. Those and others characteristics –the zones appear marked with different colours depending of data similarity and its distance to data belonging to others clusters-, allow to realize a first draft of the location of anomalous or not typical demand profiles -left bottom corner on the map-. Also, using the label map obtained by means of the criterion already explained (mm/dd), it is possible to identify the particular days assigned to the region of interest –see figure 2-.

For example, the study of figure 2 shows labels 501 and 1208, corresponding with two holidays in Spain, located in this area. Also a county holiday marked with label 1009 is located closed to the previous ones. Obviously 1231, 1225 are festive days in Christmas Season.

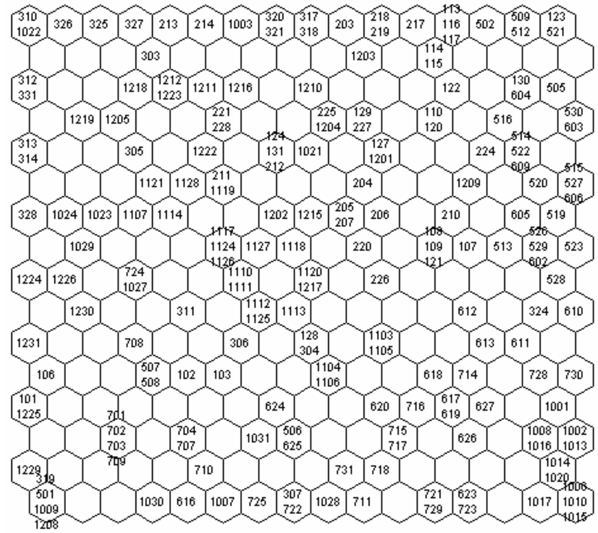

Figure 2. University Label Map (mm/dd criterion).

Finally and once the network is trained, it is possible to force it to group data fixing an upper limit of clusters. In this particular case and for a maximum of two clusters the following result was obtained -figure 3-.

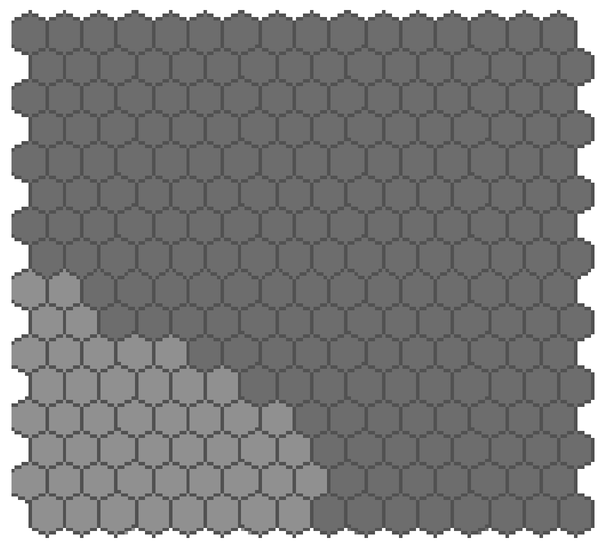

Figure 3. University Clusters Map with 2 maximum clusters.

Searching through the label map the profiles assigned to the region of minor size, it can be observed that these consumptions fit with holidays. Moreover, these distant patterns -from the typical ones- for the customer are located in the opposite corner. If the upper limit of clusters is increased -in an iterative and automatic way- it is possible to optimize the search because punctual or seasonal demand patterns are removed from standard ones.

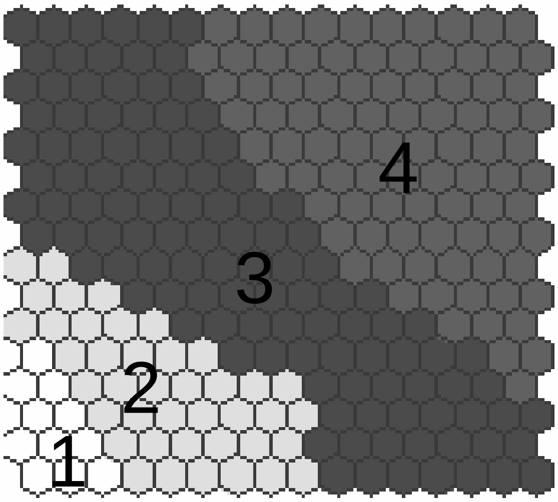

Figure 4. University Clusters Map with 10 maximum clusters.

When a maximum number of ten clusters is allowed, the four zones defined in figure 4 are found. By means of the label map and plotting the corresponding load profiles, it can be seen that the network is able to distinguish three kinds of profiles: typical consumption patterns, assigned to regions 3 and 4 -see figure 5-; profiles placed in region 1, which due to the characteristic of topographic preservation of SOM are identified as holidays –i.e., anomalous days, see figure 6-; finally profiles that denote standards of different behaviour from the usual ones lie in area 2 -see figure 7-.

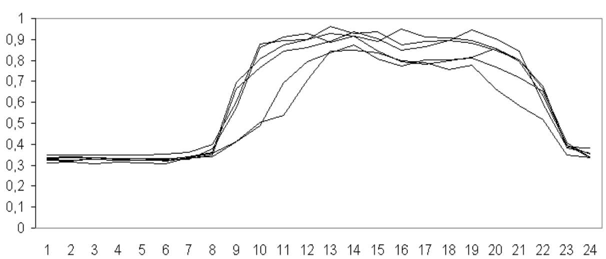

Figure 5. Typical Consumption Pattern. University

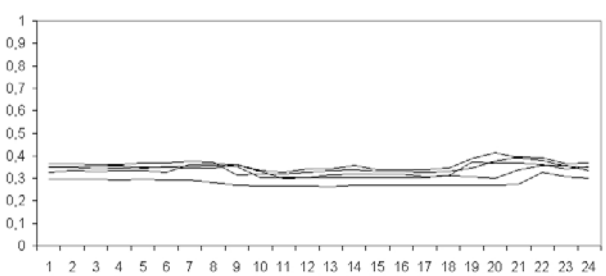

Figure 6. Holidays Consumption Pattern. University

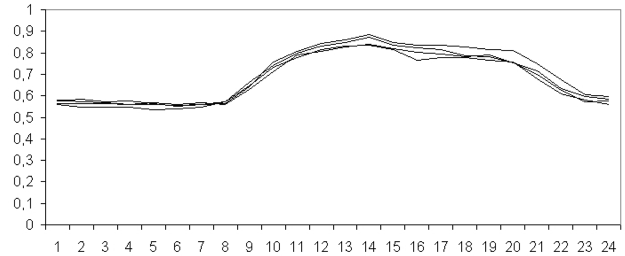

Figure 7. Change in patterns due to air conditioning. University

The previous results show as the network is not only able of identifying anomalous days with the consistent capacity of cleanliness of the database, but it presents other usefulness or applications. Thus, the network isolates consumption profiles with erroneous measures caused by failure in demand meters and will enable the utilities to manage them in a right way -to reconstruct them or to eliminate wrong data-. Moreover, it allows the study of particular behaviors of the customer. So demand patterns of the studied university identified in cluster 2 corresponds to days of July during which relative night-time demand level growths due to the continuous use of air conditioning and a day-time peak demand reduction owed to student's holidays period. Therefore, the exposed method is presented as a useful tool in the study and estimation of end use loads -air conditioning in this case- and its possible routes of management -control actions, introduction of new technologies, and efficiency measures-.

5. Customer Patterns Classification

The first purpose of application of SOM is to obtain a map for the identification and classification of electrical customers with different demand patterns that may be later employed in diverse tasks [9], [10], [11]. In the present section an example of classification is shown using data of the two customers presented at the beginning of this paper. The typical profiles of the above mentioned customers are well-known, showing features that make them notably different in their behaviors.

First, the training of the network has been done using all available information without data filtering. Simply it was followed the data conditioning process previously described using the network's characteristics shown in section 4. Label map obtained using mask 1 for Medium Industry and mask 2 for the University can be seen in figure 8. In that figure it is shown how the network is able to group the customer's profiles in the same map area separating and differentiating each one.

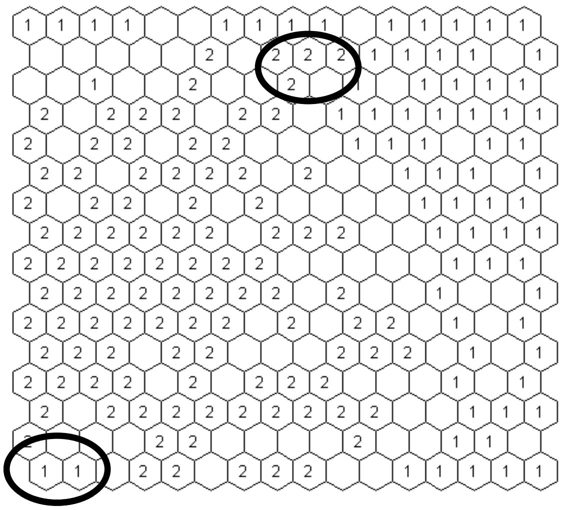

Figure 8. Customer classification without filtering

Nevertheless, it is shown how the map presents zones where separation is not clear and labels of both customers are even mixed. This fact suggests that some profiles belonging to different customers are not well suited to typical consumption patterns and therefore the network is not able to distinguish them in a right way.

For that reason it was performed a new training eliminating previously the anomalous demand profiles identified by means of the method exposed in the previous section. In this case the filtering was separately applied to both customers. The results can be seen in figure 9: the network is now able to classify correctly and besides separate with higher resolution the data corresponding to both customers.

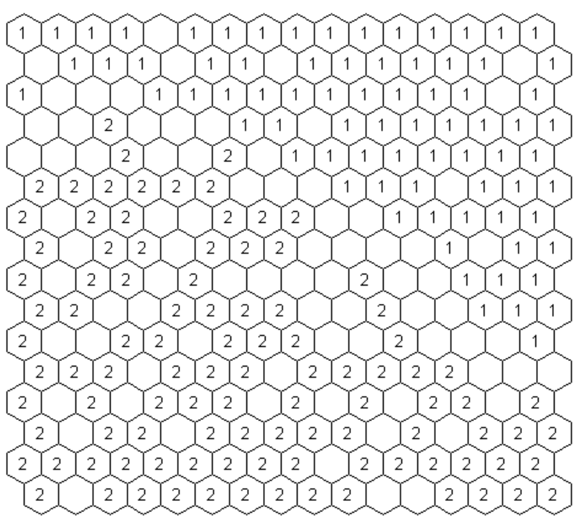

Figure 9. Customer classification with previous filtering

Typical customers' profiles classified by the SOM are can be seen in figure 10. Notice the necessity of previous database filtering due to the similarities founded between industry load profiles and university holiday profiles -a load curve with a higher load factor, see figure 6-.

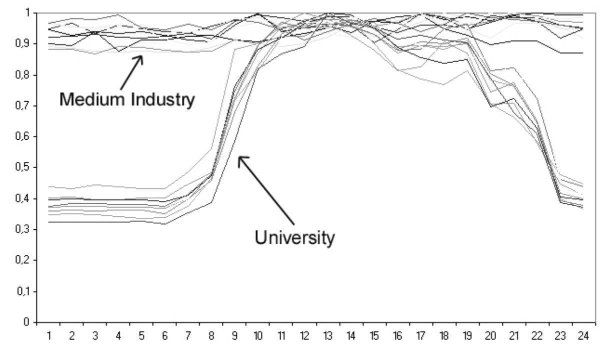

Figure 10. University and Industry load profiles

The improvement is stated if reduction in quantization error -Qe- is observed -see table II-. This error gives information of accuracy with which the map represents the samples of data. Topographic error -Te- measures the percentage of profiles that have no-closed first and second winning units. As it can be seen, the value for the second one is negligible -a small value indicates that the network does not "hesitate" in the moment of profiles assignation to cells-.

TABLE II
QUANTIZATION AND TOPOGRAPHIC ERRORS

Training	**Qe**	**Te**	**Qe Relative Reduction (%)**
With Anomalous Days	0.1520	0.0114	---
Without Anomalous Days	0.1303	0.0150	14,3

Therefore it can be concluded that through the process of previous filtering, the results of the application of SOM to a higher number of customers can lead to an excellent classification with a great usefulness.

6. Conclusions

A SOM application tool applied to distribution power system operation and management is presented in this work to achieve the segmentation of two typical electrical customers and their patterns for the energy-use on the basis of their daily load curves. In the case where possible presence of anomalous data and so uncertainty appears –such as the case of large utility databases-, this neural network provide the detection of outliers, missing information or, for example, excursions from standard pattern -due to price changes, or external factors as is the case of external temperature growth.

The method presented here can effectively help commercializers and distributors in customer segmentation and classification. This is the first step to evaluate cost-effectiveness of a great number of necessary policies in the demand-side, for example the potential of energy efficient alternatives, customer response to real price or TOU tariffs, tariff diversification, the success of dual-fuel or energy storage appliances or the possibilities of distributed generation in medium and small users.

The research activity already in study is devoted to the development of three objectives: the improvement of segmentation indices used in the SOM map –including some external parameter, such as economical activity, quality and reliability of supply, tariffs,…-, testing of segmentation in a larger database –allowing successive segmentation levels for the main clusters- and the development of new tools based in ANN to identify the percentage of a kind of customer in a distribution system. Results of these works will be reported by the authors in the near future.

7. Acknowledgements

The work described in this paper is financially supported by the Ministerio de Ciencia y Tecnología of Spain and the FEDER Funds through the research projects DPI2001-2779-C02-01 and DPI2001-2779-C02-02. Furthermore the authors would like to thank the European Union for its support.

References:

[1] California Energy Comission, *Integrated Energy Policy Report*, Docket No 02-IEP-1, Pub No 100-03-019, December 2003.

[2] C. Olaru, L. Wehenkel, Data Mining tutorial, *IEEE Computer Applications in Power, Vol. 12 , n 3* , pp. 19-25, July 1999.

[3] B. D. Pitt, D. S. Kirschen, Application of Data Mining Techniques to Load Profiling, *Proc. IEEE PICA'99*, Santa Clara, CA, May 16-21, 1999, 131-136.

[4] T. Kohonen, *Self-organisation and associative memory* (3rd ed.), (Springer-Verlag, Berlin, 1989).

[5] S. Valero, M. Ortiz, Fco. García, A. Gabaldón et al., Characterization and Identification of Electrical Customer Through the Use of SOM and Daily Load Parameters, *IEEE PSCE2004*, New York, October 10-13, 2004.

[6] G. Chicco, R. Napoli, F. Piglione, P. Postolache, M.Scutariu, C.Toader, Load Pattern-Based Classification of Electricity Customers, *IEEE Trans. On Power Systems, vol. 19, n 2, pp. 1232-1238, May 2004*

[7] G.Chicco, R.Napoli, P.Postolache, M.Scutariu and C.Toader, Customer Characterisation Options for Improving the Tariff Offer, *IEEE Transactions on Power Systems. Vol.18*, nº 1, Feb. 2003, 381-387.

[8] DeGruijter, J. J., McBartney, A. B., *A modified fuzzy k means for predictive classification*. (In: Bock, H. H.(ed) Classification and Related Methods of Data Analysis. pp. 97-104. Elsevier Science, Amsterdam, 1988).

[9] G. Chicco, R. Napoli, F. Piglione, Load Pattern Clustering for Short-Term Load Forecasting of Anomalous Days, *2001 IEEE Porto Power Tech Conference*, 10-13 September, Porto, Portugal.

[10] C. S. Chen, J. C. Hwang, Y. M. Zzeng, C. W. Huang, M. Y. Cho, Determination of Customer Load Characteristics by Load Survey System at Taipower, *IEEE Transactions on Power Delivery, Vol. 11*, No. 3, July 1996.

[11] R.Lamedica, A.Prudenzi, M.Sforna, M.Caciotta and V.Orsolini Cencelli,, A Neural Network Based Technique for Short-Term Load Forecasting of Anomalous Load Periods, *IEEE Transactions on Power Systems. Vol.11*, nº 4, 1996, pp. 1749-1756.