IEEE TRANSACTIONS ON
POWER SYSTEMS

A PUBLICATION OF THE IEEE POWER ENGINEERING SOCIETY

NOVEMBER 2006 VOLUME 21 NUMBER 4 ITPSEG (ISSN 0885-8950)

POWER SYSTEM ANALYSIS, COMPUTING, AND ECONOMICS

(Contents Continued on Page 1469)

IEEE

(Contents Continued from Front Cover)

(Contents Continued on Page 1470)

Classification, Filtering, and Identification of Electrical Customer Load Patterns Through the Use of Self-Organizing Maps

Sergio Valero Verdú, Mario Ortiz García, Carolina Senabre, Antonio Gabaldón Marín, *Member, IEEE*, and Francisco J. García Franco

Abstract—**Different methodologies are available for clustering purposes. The objective of this paper is to review the capacity of some of them and specifically to test the ability of self-organizing maps (SOMs) to filter, classify, and extract patterns from distributor, commercializer, or customer electrical demand databases. These market participants can achieve an interesting benefit through the knowledge of these patterns, for example, to evaluate the potential for distributed generation, energy efficiency, and demand-side response policies (market analysis). For simplicity, customer classification techniques usually used the historic load curves of each user. The first step in the methodology presented in this paper is anomalous data filtering: holidays, maintenance, and wrong measurements must be removed from the database. Subsequently, two different treatments (frequency and time domain) of demand data were tested to feed SOM maps and evaluate the advantages of each approach. Finally, the ability of SOM to classify new customers in different clusters is also examined. Both steps have been performed through a well-known technique: SOM maps. The results clearly show the suitability of this approach to improve data management and to easily find coherent clusters between electrical users, accounting for relevant information about weekend demand patterns.**

Index Terms—**Data mining, demand management, electrical customer segmentation, load patterns, self-organizing maps (SOMs).**

I. Introduction

THE liberalization process of the electrical market has not been as successful as was planned, due to a lot of problems that have appeared since 2000 until now: for example, the California energy crisis in 2000 or blackouts in Europe, the 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 would take into account not only supply but also demand policies: for example, efficiency gains in demand (in a long-term horizon), demand management, or price responsiveness (in short-term horizon). The effective contribution to these programs and the necessity of offering energy choices to consumers need: a detailed knowledge of customer segments, the characterization of these segments (demand behavior), end-uses "dissection" for each customer, load modeling (demand and response models), and further demand aggregation to achieve demand packages for demand-side biddings and offers in energy markets (see Fig. 1).

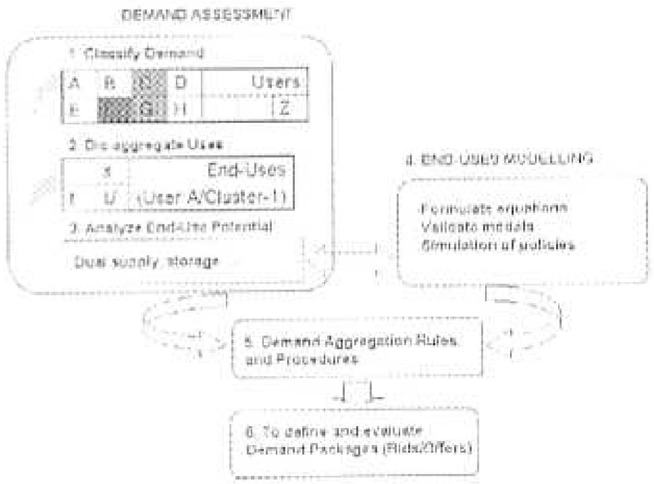

Fig. 1. Methodology to analyze, evaluate, and enhance the possibility of demand participation (DSB/DR) in electricity markets.

Besides, this deregulation and liberalization in power systems caused the necessity of new (customer and system) measurements, monitoring, and control activities. This fact has increased the amount of data stored by supply-side actors. So, this enormous quantity of available data presents a problem for utilities but also a non-negligible opportunity for distribution research. This high-dimensional data set cannot be easily modeled, and advanced tools for synthesizing structures from such information are needed.

Previous results on modeling, aggregation, and construction of energy packages were presented by the authors of [1] and [2]. The rest of this paper presents a methodology for customer segmentation and classification through the improvement and use of the data mining or knowledge discovery in databases techniques [3], [4].

Manuscript received October 4, 2005; revised May 5, 2006. This work was supported by European Union Sixth Framework Program under Project EU-DEEP SES6-CT-2003-503516. Paper no. TPWRS-00633-2005.

S. V. Verdú and M. O. García are with the Department of Electrical Engineering, Universidad Miguel Hernández, Elche, Spain (e-mail: svalero@umh.es).

C. Senabre is with the Department of Mechanics, Universidad Miguel Hernández, Elche, Spain.

A. Gabaldón Marín and F. J. G. Franco are with the Department of Electrical Engineering, Universidad Politécnica de Cartagena, Cartagena, Spain (e-mail: antonio.gabaldon@upct.es).

Digital Object Identifier 10.1109/TPWRS.2006.881133

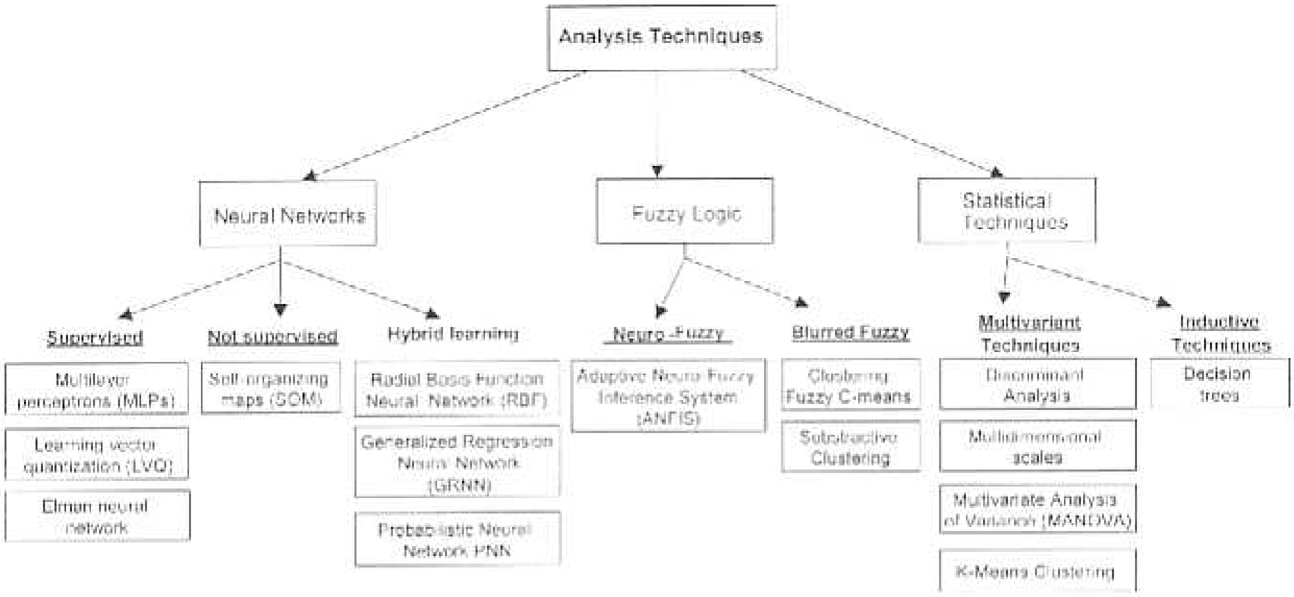

Fig. 2. Analysis techniques.

II. Review of Customer Classification Methodologies

In some data mining tutorials [3], classification methodologies are grouped in different categories according to the main task they are usually focused on: artificial intelligence techniques (neural networks and fuzzy logic), statistical techniques (linear regression and discriminant analysis), and visualization techniques (histograms, dendograms, and scatter plots). Fig. 2 shows a compendium of the techniques mentioned above and tested for this paper.

A. Techniques Review

The following paragraphs describe the characteristics of the most interesting methodologies presented in Fig. 2.

1) Artificial Neural Networks Techniques: Artificial neural networks (ANNs) try to reproduce the way the human brain acts: a highly complex, nonlinear, and parallel information processor able to perform certain computations many times faster than the most powerful digital computer available today.

Actually, ANNs find applications in such diverse fields as modeling, time-series analysis pattern recognition, and others by virtue of their ability to learn from input data with or without a teacher.

First important results in the ANN field were obtained with the simple perceptron (1958) [5] and the adaptive linear element (ADALINE) (1960), two supervised learning neural networks able to classify linearly separable sets of vectors.

Simple perceptron evolved into multilayer perceptrons (MLPs), feedforward neural networks with more than one perceptron used to solve more difficult problems.

Later, in the 1980s, Kohonen introduced the learning vector quantization (LVQ) [6] based on competitive layers in which neurons compete with each other for the right to respond to a given input vector: individual neurons learn to become feature detector cells.

Finally, Elman networks [7] are able to learn, recognize, and generate temporal patterns, as well as spatial patterns, by means of the recurrent connection feature of the network.

If the target outputs are not available, unsupervised networks must be used. In this case, the weights and biases of the network are only modified in response to inputs (so target outputs are not needed), and the algorithms classify the input patterns in a finite number of classes.

Self-organizing maps (SOMs) [6] are unsupervised networks able to learn both the distribution (as competitive layers do) and the topology of the input vectors on which they are trained. Consequently, excellent clustering results are obtained. In addition, an easy evaluation of the result is possible through the graphical representation on maps whose different labels (customers or vectors identifiers) can be grouped by visual inspection. Applying some index functions, it is possible to obtain an optimum clustering, but some "supervision" is necessary to filter the results of the maps (i.e., the operator selects the maximum number of clusters). More detailed information is presented in Section III.

The main features of the supervised and unsupervised techniques discussed above can be consulted in Table I.

Some methodologies in Fig. 2 appear as "hybrid learning" techniques. A hybrid method for learning encompasses two phases: the first is a not supervised one for the determination of clusters center, and the second is a supervised phase, for the weights and thresholds determination [8].

Three different techniques are presented: radial basis networks [9], generalized regression neural networks (GRNN) [10], and probabilistic neural networks (PNN) [9], [11]. The GRNN and PNN have a disadvantage: they perform the operations slower than other kinds of networks [12], [13].

B) Fuzzy Logic Techniques: Another interesting possibility, for clustering purposes, is the use of fuzzy methods: ANFIS [14], fuzzy C-means, originally introduced by Bezdek in 1981