Electrical Power and Energy Systems 53 (2013) 824–831

Contents lists available at SciVerse ScienceDirect

Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes

A new method for pattern recognition in load profiles to support decision-making in the management of the electric sector

Adonias M.S. Ferreira [a,*], Carlos A.M.T. Cavalcante [a], Cristiano H.O. Fontes [a], Jorge E.S. Marambio [b,1]

[a] Program of Industrial Engineering, Polytechnic School, Federal University of Bahia, Rua Professor Aristides Novis, 2, Federação, CEP 40210-630 Salvador, BA, Brazil
[b] Norsul Engenharia LTDA, Av. Tancredo Neves 1632, S1802, Edf. Salvador Trade Center, Torre Sul, Caminho das Árvores, CEP 41820 020 Salvador, BA, Brazil

ARTICLE INFO

Article history:
Received 25 July 2012
Received in revised form 29 May 2013
Accepted 1 June 2013

Keywords:
Load profiles
Clustering
Pattern recognition
Electric sector

ABSTRACT

This work presents a method for the selection, typification and clustering of load curves (STCL) capable of recognizing consumption patterns in the electricity sector. The algorithm comprises four steps that extract essential features from the load curve of residential users with an emphasis on their seasonal and temporal profile, among others. The method was successfully implemented and tested in the context of an energy efficiency program carried out by the Energy Company of Maranhão (Brazil). This program involved the replacement of refrigerators in low-income consumers' homes in several towns located within the state of Maranhão (Brazil). The results were compared with a well known time series clustering method already established in the literature, Fuzzy CMeans (FCM). The results reveal the viability of the STCL method in recognizing patterns and in generating conclusions coherent with the reality of the electricity sector. The proposed method is also useful to support decision-making at management level.

1. Introduction

Multivariate analysis is a powerful tool for knowledge extraction especially when applying pattern recognition techniques based on data. In this context, the analysis of energy consumption data measured in homes can identify opportunities for improvement in the load factor [1] and energy efficiency of the distribution system through specific actions by the customer [2]. Methods of Data Mining (DM) that can extract useful information from data can be used to develop decision-making tools so as to improve production systems and management technology [3–6].

Some works present the use of clustering and load curve typification (pattern recognition) methods in the electric power sector. Gerbec et al. [7] performed a load curve typification using a hierarchical clustering method highlighting the advantage of this method in choosing the appropriate number of groups. The non-hierarchical method [8] emphasizes the minimization of internal variance within a cluster and also the reduction of similarity between different groups. Geminagnani et al. [9] combined the hierarchical and non-hierarchical clustering methods to improve clustering efficiency in the recognition of different consumption patterns at the same level of tension. Zalewski [10] used fuzzy logic for clustering and load curve typification. The author performed the clustering of load profiles in order to classify substations into homogeneous groups according to consumption peak. Nizar et al. [11] combined two methods, namely, Feature Selection and Knowledge Discovery in Databases (KDD) [12,13], to obtain better patterns of load demand in a distribution system. A recent study about knowledge extraction from electric power consumer data [10] presents an overall analysis and prediction of energy consumption trends (Incremental Summarization and Pattern Characterization – ISPC).

Some studies compare the performance of various methods of typification and conclude that the fuzzy C-Means (FCM) provides the best level of cohesion and discrimination of the problems associated with clustering in load curves. From this very point of view some authors have recently highlighted the FCM method in applications involving pattern recognition (typification) in load curves [14–16].

This study proposes a new method of selection, typification, and load curve clustering (STCL) based on a systematic extraction of features. This method is capable of identifying a greater diversity in demand patterns and also represents a potential tool for the improvement of the decision-making process through better classification of heterogeneous consumer profiles in the electric power sector. The case study analyzed is an energy efficiency program [17] carried out by the Electric Company of Maranhão (Brazil), that considers, among others, the analysis of the impact of replacing refrigerators in low-income consumers' homes distributed in several towns located within the state of Maranhão (Brazil). The proposed method incorporates multiple criteria in the clustering

* Corresponding author. Tel.: +55 (71) 3203 9806; fax: +55 (71) 3203 9802.
E-mail addresses: adoniasmagdiel@ufba.br (A.M.S. Ferreira), arthurtc@ufba.br (C.A.M.T. Cavalcante), cfontes@ufba.br (C.H.O. Fontes), contato@norsulengenharia.com.br (J.E.S. Marambio).
[1] Tel./fax: +55 (71) 3342 7013.

0142-0615/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ijepes.2013.06.001

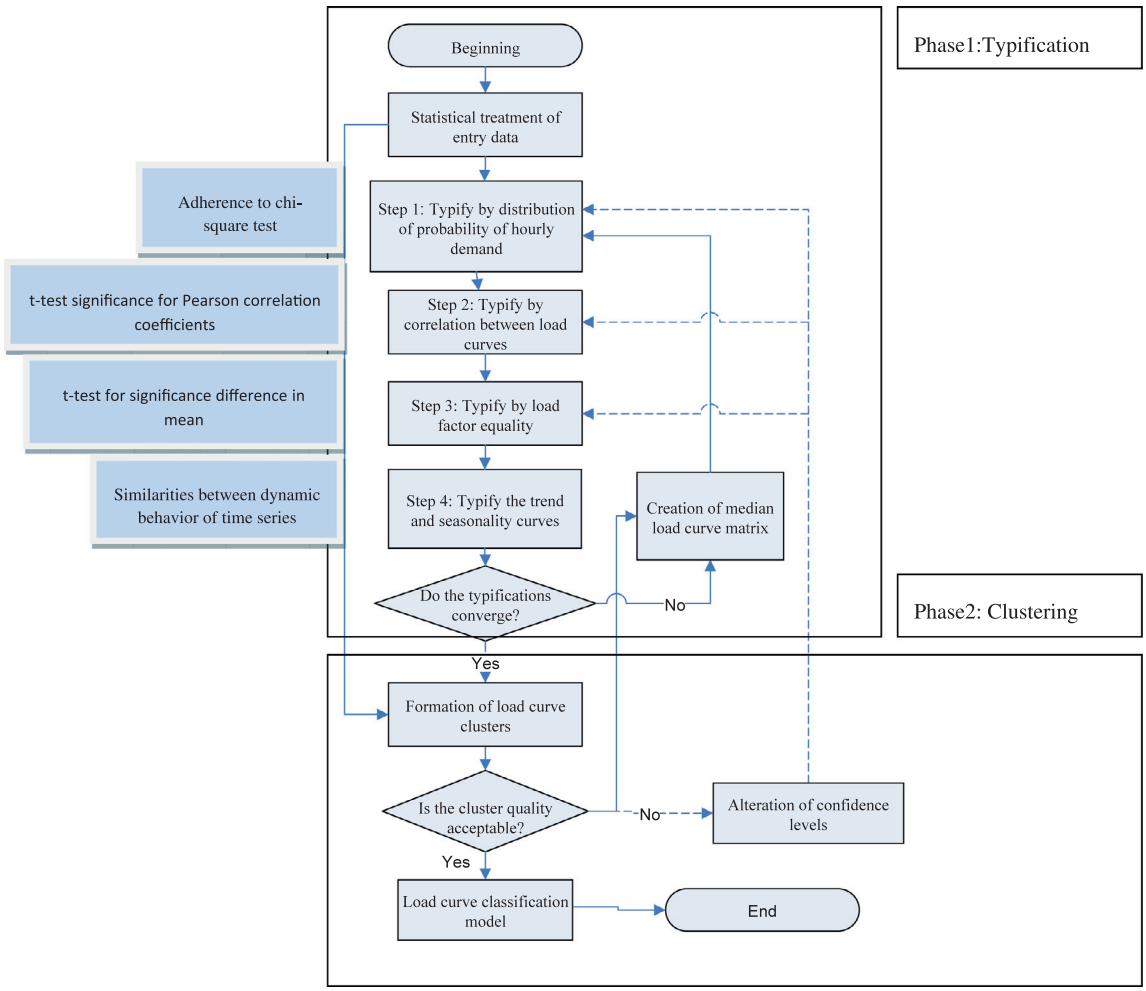

Fig. 1. The STCL method.

and typification of load curves unlike traditional approaches that essentially use the criterion of distance between load curves for cluster recognition. Section 2 presents the STCL method and the evaluation metrics adopted. Section 3 presents the case study and results obtained from the application of FCM and STCL methods demonstrating the ability and superiority of the latter in describing the problem.

2. The STCL method

The STCL method (Fig. 1) comprises two phases. The first carries out pattern recognition through successive iterations. The first iteration performs the clustering of the whole sample (load curves from the database) based on specific features associated with the consumption profile, and some clusters of load curves are obtained. The subsequent iterations consider only the medians (patterns) of each group generated and verify the similarity between these medians based on the same statistical tests considered in the first iteration such that some patterns may be collapsed. Thus, at the end of the first phase (after convergence), patterns or types associated with load curves are recognized. The second phase defines the final groups associating each load curve (database) to one of the patterns recognized in the first phase.

Initially, each curve is normalized within the interval [0; 1] dividing the hourly measurements by the peak demand of each. The dimensionless consumption quantified in this way is called power per unit (pu) [16].

Table 1 presents the criteria considered in the similarity analysis performed in each step of the first phase together with the statistical test applied. These criteria were established according to the requirements and indicators practiced in the electric energy distribution sector [10,18–21].

The three features (three stages) presented in Table 1 are applied successively. In the first iteration, the clusters are formed based on similarity between the load curves and the curve with the highest average power consumption (reference curve). After the first iteration, the method assumes that the median of each group keeps its own features and the same tests are successively applied considering the medians (patterns). The existence of a similarity between medians according to the statistical tests implies the union of groups and new medians are obtained.

The LF presented in Table 1 is the ratio between the average and maximum demands of a load curve. The LF is an evaluation index for the rational use of electric power by the consumer [20]. From

Table 1
Statistical tests used in sequence in the process of typing.

Clustering criterion	Statistical test
Distribution of probability of hourly demand	Chi-square for goodness of fit [16]
Selection by correlation level between load curve	Independent two-sample *t*-test significance for Pearson correlation coefficients [17]
Selection by mean consumption or load factor	Independent two-sample *t*-test for significance difference in mean [17]