Clustering Electric Load Curves: The Brazilian experience

José Francisco Moreira Pessanha[1], Luiz da Costa Laurencel[2]

[1] Centro de Pesquisas de Energia Elétrica (Cepel), francisc@cepel.br
[2] Universidade do Estado do Rio de Janeiro (UERJ/UFF), getlcl@vm.uff.br

Abstract This contribution presents a brief history of softwares for identifying of typical daily load profiles developed in the Brazilian electric power sector since the eighties. In order to tell this history we present the features of three softwares that have been used by the Brazilian electric distribution utilities, all of then with data mining techniques for clustering.

Keywords: cluster analysis, electric load curves, Brazilian power sector

1 Introduction

The Brazilian electric power sector adopts tariffs based on marginal cost pricing since 1980 [1]. The electricity tariffs are calculated by a methodology, whose origin is the Electricité de France (EDF) and the French 'marginaliste' economists like Allais and Boiteux [2]. In this methodology, an important step is the identification of a few typical daily load profiles from a set of electric load curves measured on a sample of customers. These profiles represent patterns of energy use at different class of customers e.g. residential, commercial, industrial, rural, public lighting, public administration etc. The input data required to the modelling of typical load profiles is a sample of electric load curve measurements. In general, the load curves measures cover a period of two weeks (15 days), where the demand is recorded every 15 minutes by recording meters installed at each consumer point in the sample. In order to reduce the data dimensionality three representative days must be selected from each load curve measurement file: a workday, a Saturday and a Sunday. Each representative daily curve is a vector with 96 points (demand is recorded every 15 minutes). The standard way to identify the typical load profile from a sample of load curves is to perform a clustering of the representative workday load curves classified in a same customer class. The centroid of each cluster defines a typical load profile.

2 Softwares

In the SNACC, the first program developed in the Brazilian power sector in 1982, were adopted the methods of cluster analysis programmed in two computational routines in FORTRAN brought from France: NUDYC and DESCR2. These routines are executed sequentially: first the typical workday load curves are clustered by the nuées dynamique" [1,3,4], a non-hierarchical method programmed in the NUDYC routine, then the clusters centroid vectors are clustered by the Ward method [1,3] (hierarchical method) programmed in the DESCR2 routine. Despite these sophisticated routines to identify the typical load profiles, the SNACC has not a friendly user interface. The program does not

show graphs of load curve measures and typical load profiles, an important output for the load curve analysis. This deficiency was the main critical step for the computation of distribution tariff. In order to overcome this deficiency, in 1998 the Brazilian Electric Power Research Center (Cepel) developed the TARDIST system [5] for calculating the distribution tariffs based on marginal cost. This software also has a module to build typical load profiles with a friendly user interface, where, for example, one can choice the three representative days of each load curve measurements based on a graphical output. The software employs only the Ward method in order to derive typical load profiles from a set of workday load curves. The graphical interface shows the workday load curves classified in each cluster and the resulting centroid, i.e., the typical load profile. Also are presented the within and between group sum of squares (in absolute and relative values) for different aggregation level. These statistics and the graphical outputs help us determine the number of typical load profiles. More recently the Cepel developed the ANATIPO, a new software to identify typical load profiles from a sample of load curves measurements. This software incorporates several advances in graphical interface techniques, for example, facilitate the selection of the three representative load curves from a file measurement and also allow move workday load curves from a cluster to another. The software offers three methods of cluster analysis: k-Means, Ward and Fuzzy cluster method (FCM). The typical load profiles obtained by the ANATIPO are organized in electronic worksheets ready to attach in the tariff proposal to be send to ANEEL (the Brazilian regulatory agency). Following the international trend [6,7], nowadays in Brazil the researches on the area have been concentrated on the application of the Self-Organizing Map, a unsupervised neural network.

References

[1] BRASIL, Ministério das Minas e Energia, DNAEE, Eletrobrás, Empresas Concessionárias de Energia Elétrica, Nova Tarifa de Energia Elétrica: metodologia e aplicação, DNAEE, Brasília, 1985.

[2] Boiteux, M. La tarification dês demandes en pointe: application de la théorie de la vente au coût marginal, Revue générale de l'electricité, 1949.

[3] Lebart, L.; Piron, M.; Morineau, A. Statistique exploratoire multidimensionnelle, 3e édition, DUNOD, Paris, 2000.

[4] Bouroche J.M. et Saporta G., L'analyse dês données, PUF, 9e édition, Paris, 2005.

[5] Pessanha, J.F.M., Huang, J.L.C., Pereira, L.A.C., Passos Júnior, R., Castellani, V.L.O. Metodologia e sistema computacional para cálculo das tarifas de uso dos sistemas de distribuição, XXXVI SBPO, São João del Rey - MG,2004.

[6] Debrégeas, A., Hébrail, G. Interactive interpretation of Kohonen maps applied to curves, International Conference on Knowlodge Discovery and Data Mining, New York, August, 1998.

[7] Hébrail, G. Practical data mining in a large utility company, Revue Questiio (Quaderns d'Estadistica i Investigacio Operativa), Vol.25, N.3, pp.509-520, 2001.