CLUSTERING TECHNIQUES IN LOAD PROFILE ANALYSE

Elena Crenguta BOBRIC
"Stefan cel Mare" University of Suceava
str. Universitatii nr.13, RO-720225 Suceava
crengutab@eed.usv.ro

Abstract. *In distributions network existing little information about loading level of transformer of substations. Feeders and loads are not monitoring usually. Therefore, in each moment existing one incertitude degree about buses loads and accordingly and about load level of network, voltage level by buses and power loses. Accordingly, in distribution electric network is important to estimate the load profiles of buses, as one alternative at determination of energy consumption from effectuated measurement in busses network. The paper presents one analyse at load profile for a distribution substation utilising clustering techniques*
Keywords: *load profile, clustering techniques, characteristic day, consumer tips*

Introduction

Generally, the load of busses distribution network are in a continuous development in time, pursuant to of a utilisation number more and more big of electric receivers, quotients and by reason of the development of logistics for population.
The models for electric loads determinations are different, depending on the networking tips: urban, rural and industrial.
Obtain of information referring to load from busses networks is achieved through: load teletransmission from busses networks, use the half-automatic local devices of measure the electric loads, use the electronic modern systems for data acquisitions and use primary information concerning electric loads from busses.
Obtain last categories of information's require the effectuation of measurements of the electric load of the consumers of certain in characteristic periods: winter, the summer, working day and weekend day, thereto is remake result of the effectuated measurements.

Temporally variation of the electric load reflected the graphic of daily, seasonal and annual, those carry to remake a how much number of real dates of consumption of electric energy.

The electric distribution network have a big number of load's busses even if taken the in consider merely the busses which in place the substation. The consumers connect in the network busses are and they very numerous, heterogeneous as the absorbed powers, using his technologists the social behaviors, enforcing the particular loops of thing.
This difficulty is eliminating if in under consideration of regimes ale these network is utilized the graph of daily for each bus, in the characteristic regimes (winter and the summer, working day and weekend day). These constitute on the strength of load graph tip of the consumers.

Load profile based modelling

Distribution system planners traditionally approach load modeling by estimating maximum demand values of active and reactive power, in conjunction with coincidence factors at various network levels. Although this approach has been adequate, it has several drawbacks:

- the typical load behavior during off-peak periods is unknown.
- there are inherent inaccuracies due to the use the uncertainty data and due in that the measurements of network can't be completed.

- energy calculations, especially losses of active and reactive power, are not very accurate.
- voltages at various network busses are unknown.

Using modern load-flow analysis software it is now possible to perform multiple calculations over any time period using load profiles. The advantages of performing load profile based studies include:

- load factors do not need to be estimated since load profiles inherently possess these properties.
- accurate energy and energy loss calculations can be estimated at any point in the network.
- network loadings and voltages are known for all time intervals.
- transformer tap settings can be optimized for both peak and off-peak periods.
- line drop compensation can be accurately modeled.
- the effect of additional load and load growth is accurately modeled as coincidence is naturally accounted for.

Therefore, power system load has gained more attention in the last years. It is still considered as one of the most uncertain and difficult components to model due to the large number of diverse load components, to its high distribution, variable composition with time of day and week, weather and through time, and also because of lack of precise information on the composition of the load. Different utilities are available for load forecasting purposes but also new techniques for the determination of the load characteristics from measured composition data have been developed. The result of these new techniques will lead to a better understanding of the load dynamics and therefore to an improved load representation, making it possible to decrease uncertainty margins, resulting in a positive impact on both economy and reliability of the system operation. Moreover the combination of an accurate load model and a real-time monitoring application will bring up new competitive possibilities for the distribution networks.

The loading curve representation by their mean and standard deviation curves is useful for engineering calculation and statistical analysis. A performance criterion can also be established based on probabilistic value. In urban and rural distribution network, the active and reactive loads are submitted, in every moment of normally distribution law. The calculus expressions for there two characteristic sizes: mean (1) and standard deviation (2).

$$\overline{P} = \frac{\sum_{i=1}^{n} P_i}{n}, \overline{Q} = \frac{\sum_{i=1}^{n} Q_i}{n} \tag{1}$$

$$\sigma_P = \sqrt{\frac{\sum_{i=1}^{n} P_i^2}{n} - \overline{P}^2}, \sigma_Q = \sqrt{\frac{\sum_{i=1}^{n} Q_i^2}{n} - \overline{Q}^2} \tag{2}$$

where:

P_i, Q_i - active respectively reactive load, measurement in *i* moment

$\overline{P}, \overline{Q}$ - active respectively reactive mean load

Clustering methods

Data analysis underlies many computing applications, either in a design phase or as part of their on-line operations. Cluster analysis is the organization of a collection of patterns (usually represented as a vector of measurements or a point in a multidimensional space) into clusters based on similarity. Intuitively, patterns within a valid cluster are more similar to each other than they are to a pattern belonging to a different cluster.

Clustering is useful in several exploratory pattern-analysis, grouping, decision-making, and machine-learning situations, including data mining, document retrieval, image segmentation, and pattern classification. However, in many such problems, there is little prior information (e.g., statistical models) available about the data, and the decision-maker must make as few assumptions about the data as possible. It is under these restrictions that clustering methodology is particularly appropriate for the exploration of interrelationships among the data points to make an assessment (perhaps preliminary) of their structure. The term "clustering" is used in several research communities to describe methods for grouping of unlabeled data. These communities have different terminologies and assumptions for the components of the clustering process and the contexts in which clustering are used.

Clustering is the technique of grouping rows together that shares similar values across a number of variables. It is a wonderful exploratory technique to help you understand the clumping structure of your data. There are two major methods of clustering: hierarchical clustering and k-means clustering.

In hierarchical clustering the data are not partitioned into a particular cluster in a single step. Instead, a series of partitions takes place, which may run from a single cluster containing all objects to n clusters each containing a single object. Hierarchical clustering is subdivided into *agglomerative* methods, which proceed by series of fusions of the n objects into groups, and *divisive* methods, which separate n objects successively into finer groupings. Agglomerative techniques are more commonly used. Hierarchical clustering may be represented by a two dimensional diagram known as dendrogram which illustrates the fusion or divisions made at each successive stage of analysis.

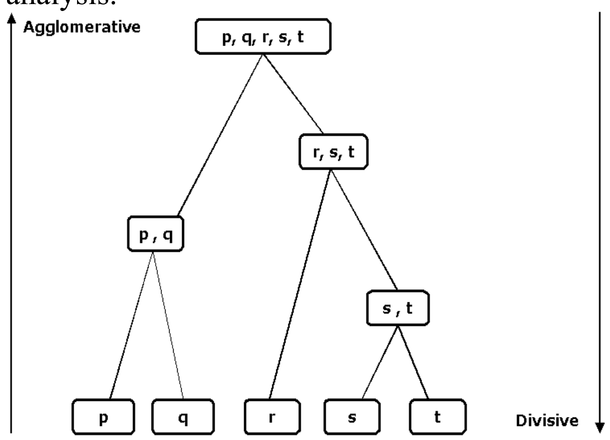

Figure 1. Clustering dendogram

Hierachical clustering is appropriate for small tables (until several hundred rows). You can choose the number of clusters you like after the tree is built. Several agglomerative techniques are single linkage clustering, complete linkage clustering, average linkage clustering, centroid method and Ward's hierarchical clustering method.

Differences between methods arise because of the different ways of defining distance (or similarity) between clusters.

In the centroid method, method utilizing in this analysis, the distance between two clusters is defined as the squared Euclidean distance between their means. The centroid method is more robust to outliers than most other hierachical methods but in other respects may not perform as well as Ward's method or average linkage:

$$D_{KL} = \left\| \overline{X}_K - \overline{X}_L \right\|^2 \tag{3}$$

K-means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori. The main idea is to define k centroids, one for each cluster. These centroids shoud be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed and an early groupage is done. At this point we need to re-calculate k new centroids as barycenters of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated. As a result of this loop we may notice that the k centroids change their location step by step until no more changes are done. In other words centroids do not move any more. Finally, these algorithms aim at minimizing an *objective function*, in this case a squared error function. The objective function:

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2 \qquad (4)$$

where $\left\| x_i^{(j)} - c_j \right\|^2$ is a chosen distance measure between a data point $x_i^{(j)}$ and the cluster centre c_j.

J is an indicator of the distance of the *n* data points from their respective cluster centres. K-means clustering is suitable for larger tables, up to hundreds of thousands of rows.

Case study

This paper proposed the analysis of load profile recording for two mounts period, of part from 20 kV of transformer by 16 MVA of the distribution substation 110/20 kV.

The load diagram is constructed using the register of watt-hour meter. The time interval of sampling load curve data is 15 minutes. Thus, the load profile is represented by 96 load values throughout of the day.

The consumption is preponderant formatted by the rural consumers. Addition this, exist the little industrial consumers and the town consumers, but these consumption not representing moreover 10% by the monthly electric energy consumption.

The analyses are effectuated on 58 days, registration by 20 February – 20 April period.

In figure 2 are represented the mean load profile for characteristic days from March.

Every measurement effectuated must be elaborated through these arranged and normalized.

$$z_{ij} = \frac{x_{ij}}{\sum_{i=1}^{24} x_{ij}} \qquad (5)$$

where :

z_{ij} - normalised value for active power

x_{ij} - measured value for active power

$\sum_{i=1}^{24} x_{ij}$ - daily energy consumption

Energy consumption is factor using for normalized.

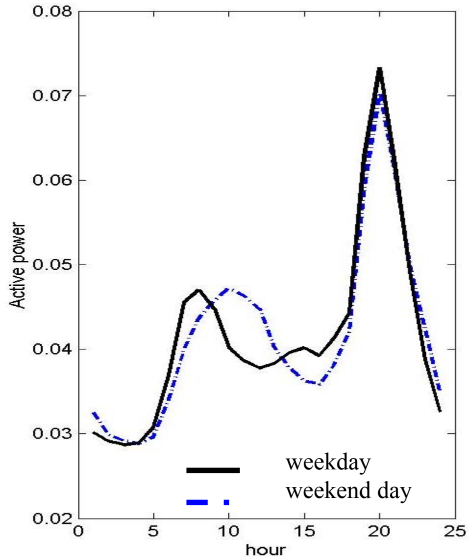

Figure 2. Load profiles for characteristic days

Classification of the load profiles through the visual comparison is subjective and impractical. The cluster analysis method was applied to solve that problem. The clustered process is making progressive into coherent and representative cluster.

The centroid method is using as clustering method. This method is applied for the dates formatted by load profile for active power for these 58 days.

In figure 3 is represented the dendogram for load profile clustering.

If we analyse this dendogram it observed that have result three cluster, who, several excepting, have be realized by profiles load accordingly whit three time periods: 21-25 February, 1 – 20 Mars and respectively 21 Mars – 19 April.

Every cluster is formatted by two subgroups corresponding the characteristic days: weekdays and weekend days. As well, existing several days they load profile for active power not enclosed in any by three clusters.

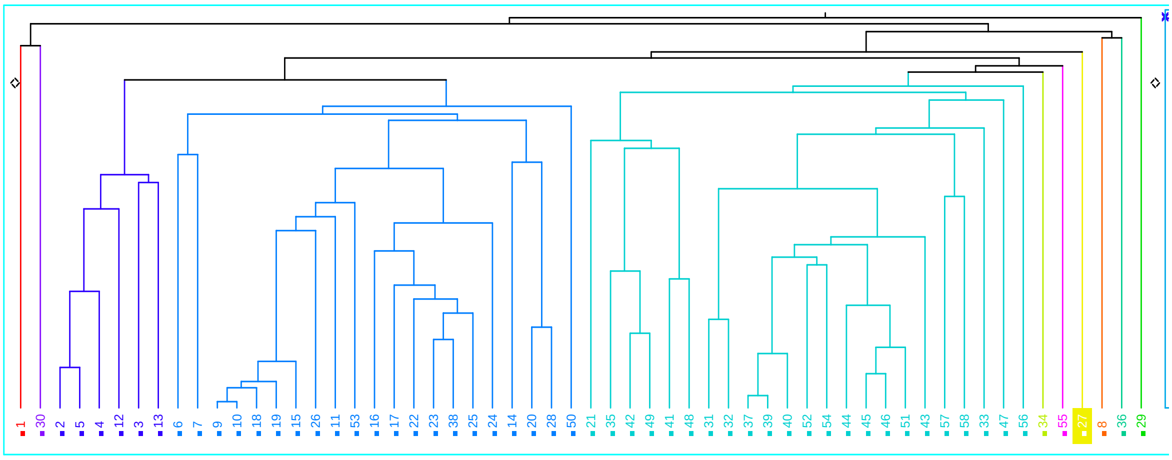

Figure 3. The dendogram for cluster the load profiles

It possible to make observation that a majored influence about consumption structure has the atmospheric conditions, specially, the environment temperature. If is analyzing the load profile for neclustering days, it is observed that standard deviation against mean load profile is big, beside that for one clustering day. This thing can be observed in figure 4.

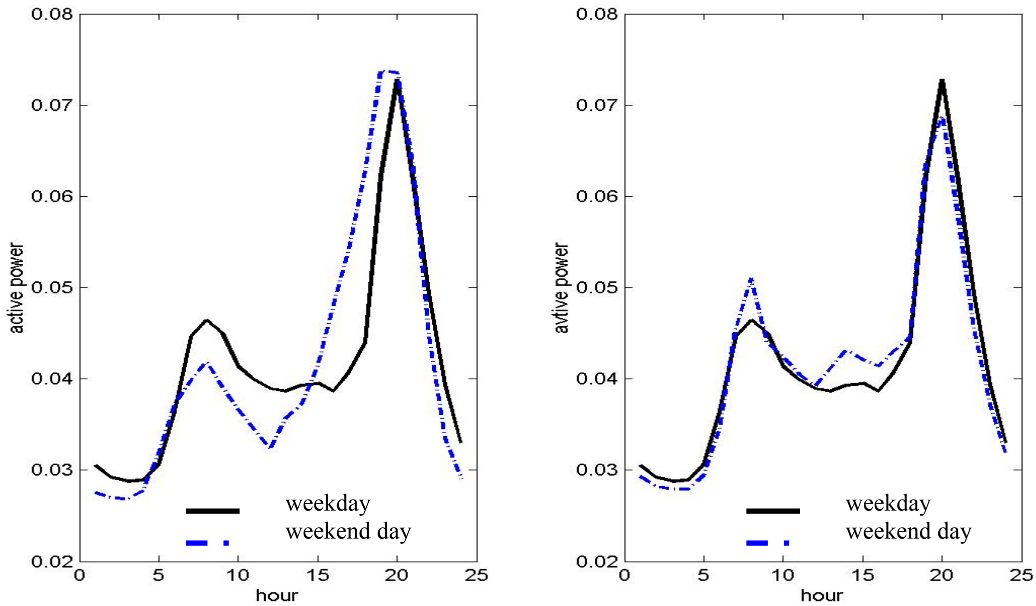

Figure 4. Mean profile and load profile out of cluster respectively one load profile from cluster

If it eliminated the load profile for stain's day without clusters and it effectuated a new grouping. It observes that for each forenamed period is clustering separate in weekdays and weekend days (figure 5).

Also, it observes that existing 3 exceptions from clustering technique. Analyzing the temperature can't say that average of day is much different from other days.

Is possible as load emptiness between 15-18 hours for load profile, of the last analyzed period, is overcast by demand load by consumer who has the reserved supply from this substation.

This analysis to be continued if allowance by the other factors that influenced the consumption structure and/or allowance by typical load profile for the consumer suppliers from this distribution substation.

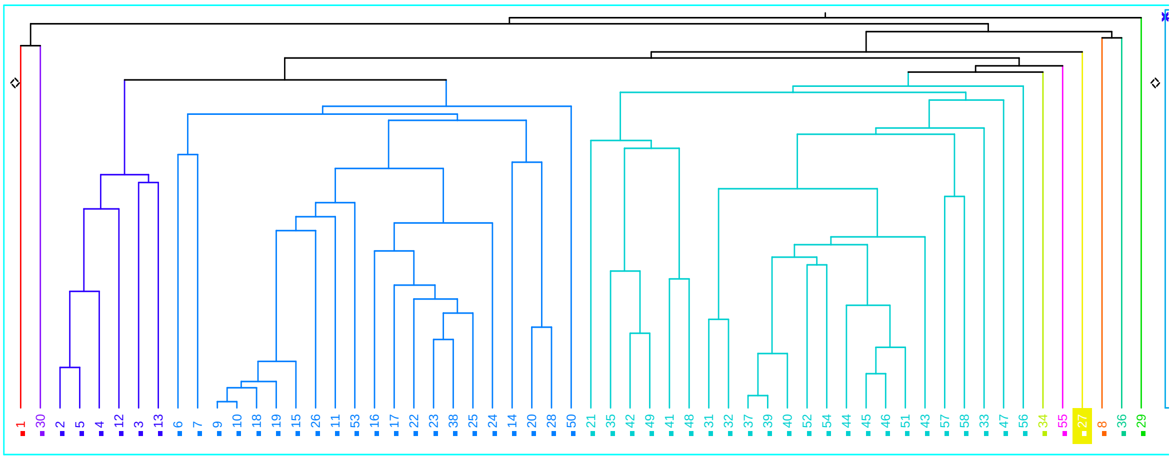

Figure 5. The dendogram for cluster the load profiles

Conclusions

In this paper is proposed one comparative analyzed method for daily load profiles for one distribution substation, for a period from two month, between 21February – 19 April.
Classification of the load profiles through the visual comparison is subjective and impractical. The cluster analysis method was applied to solve that problem. The clustered process is making progressive into coherent and representative cluster.
For this analysis was using the hierarchical clustering techniques, exactly the centroid method.
The temperature is a influenced factor for the demand power of consumer, fact that is evidentially from the dendogram obtained behind the clustering process.
The dendodram evidenced the clustering of the load profile by the characteristic days, week days and weekend days, for there three analyzed periods.
The results demonstrate the ability of clustering techniques for classification of load profile and for comparative analyses of consumption.

Acknowledgement

I would like to express my gratitude to prof.dr.ing. Gheorghe Cartina for his guidance and support in my scientific activity.

References

[1] Cârţină, Gh., Grigoraş, Gh., (2002) *Tehnici moderne de optimizare. Aplicații în energetică (Modern Optimization Techniques. Applications in Power Systems),* Iaşi, Casa de Editură VENUS.
[2] Cârţină, Gh., Grigoraş, Gh., (2004) *Tehnici de Inteligenţă Artificială în Electroenergetică*, Iaşi, Editură SETS.
[3] Cârţină Gh., Song Y.H., Grigoraş Gh., (2003) *Optimal Operation and Planning of Power Szstems*, Iasi, Casa de Editură VENUS.
[4] *JMP Statistics and Graphics Gu*ide, Version 3, SAS Institute Inc., Cary, NC, USA, 1999.
[5] Clustering: An introduction, http://www.elet.polimi.it/upload/matteucc/Clustering/tutorial_html/
[6] Gasperic, S., Gerbek D., Gubina F., *Determination of the Consumers' Load Profiles,* http://www.telmark.org/2002Sep/Gasperic.pdf
[7] Chicco G., Napoli R., Piglione F., Postolache P., Scutariu M., Toader C., *A Review of Concepts and Techniques for Emergent Customer Categorisation*, http://www.telmark.org/2002Sep/Chicco.pdf