[Downloaded from www.aece.ro on Sunday, February 11, 2018 at 22:44:33 (UTC) by 125.70.148.55. Redistribution subject to AECE license or copyright.]

Clustering Techniques in Load Profile Analysis for Distribution Stations

Elena C. BOBRIC[1], Gheorghe CARTINA[2], Gheorghe GRIGORAŞ[2]
[1]*Stefan cel Mare University of Suceava*
str.Universitatii nr.13, RO-720229 Suceava,Romania
[2] *Gheorghe Asachi Technical University of Iaşi*
Bd. D. Mangeron nr.67, RO-700050 Iaşi, Romania
crengutab@eed.usv.ro

***Abstract*—The demand characteristic is the most important one in analyzing customer information. In a distribution network, there is in any moment certain degree of uncertainty about busses loads, and consequently, about load level of network, busses voltage level, and power losses. Therefore, it is very important to estimate first of all the load profiles of buses, using available data (measurements effectuated in distribution stations). The results obtained for various distribution stations demonstrate the effectiveness of the present method in overcoming the difficulties encountered in optimal planning and operation of distribution networks.**

***Index Terms*—load profile, clustering techniques, data flow analysis, power consumption, distribution station**

I. INTRODUCTION

Electric distribution networks have a large number of load busses, even if we were to take into consideration only the busses with substations. The consumers connected in the network busses are also very numerous, heterogeneous as the absorbed powers, using their technologists, social behaviors, enforcing the particular loop of thing.

In addition, the multiple participants in the electricity market need new business strategies for providing value added services to customer. They need, therefore, accurate customer information about the electricity demand. The Demand characteristic is the most important one for analyzing customer information.

These difficulties are eliminated if for the distribution networks analysis a daily load curve is used for each bus, within characteristic regimes (winter and summer, working day and weekend day).

The models for electric loads determinations are different, depending on the networking tips: urban, rural or industrial. Variation in time of the electric load reflects the graphic of daily, seasonal and annual load, which indicate the real electric energy consumption.

In this paper, load profile data, which can be collected by means of the automatic meter reading system, are analyzed in order to get demand patterns of customers. The load profile data include electricity demand at a 15 minutes interval. An algorithm for clustering similar patterns is developed using the load profile data. As a result of the classification, representative curves for the same groups are generated. The demand characteristics of the groups are further discussed.

II. GENERAL CONSIDERATIONS

Since the storage of electric energy on a large scale is not possible, the main role of the power network is to transport the demanded energy to consumers. Therefore, it is very important to study and analyze the evolution of the load in order to operate and design the power network. All the other decisions are based on the consumed energy such as the load forecast, the voltage control, determination of the peak load for various types of consumer, calculations of the power losses or power losses estimations, proper tariff design, etc.

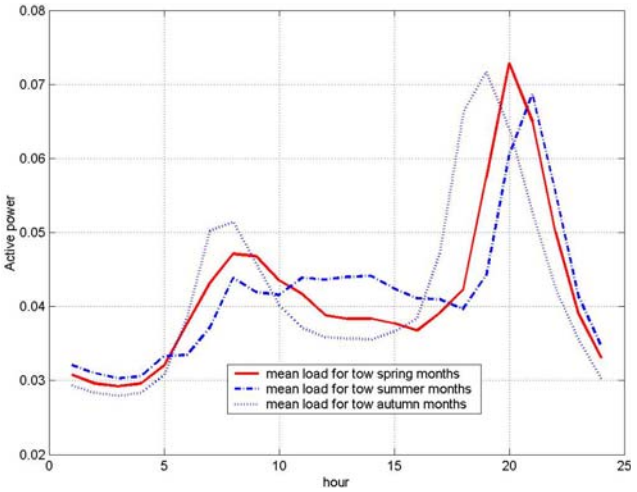


Figure 1. Load profile for different seasons.

The main causes generating load modifications are:

- weather conditions: the season, the daily temperatures, the speed of the wind, etc; (Fig. 1)
- demographic factors: the growth rate of the population, the number of the inhabitants in a certain area, the birth rate, etc;
- economic factors: the gross national product, the labor productivity, the economy development rate, the level of life quality and a very important element: the price of energy.

The evolution in time of these parameters has a strong random character. At a certain moment, the more or less accidental realization of these parameters directly influences the load and its variation change tendency influences in a decisive way the load curves.

Digital Object Identifier 10.4316/AECE.2009.01011

[Downloaded from www.aece.ro on Sunday, February 11, 2018 at 22:44:33 (UTC) by 125.70.148.55. Redistribution subject to AECE license or copyright.]

The load curve represents the power variation in terms of the determinant parameter. If the parameter taken into consideration is the time (t), the curve can be divided into several components that induce the load profile:

1. The trend (T) is the main load component, establishing the main load variation form
2. The cyclic component (C). It is due to some slow-varying causes such as the correlation supply-demand, which lasts more than a year
3. The seasonal component (S) is caused by certain parameters, which represent seasonal fluctuations. The variation period of this component lasts only a few months and it is almost the same for all years.
4. The random component (ε) is due to accidental causes that have not been mentioned above.

Therefore, the load is due to the summing up of the above- mentioned components:

$$P(t) = T(t) + C(t) + S(t) + \varepsilon(t) \quad (1)$$

The shape of the load profiles usually describes a daily and weekly periodicity. However, the load profile for tomorrow or for the next week is not just a simple copy of the load profile from today or from this week. Instead, the load profile is slightly modified from day to day and from week to week, to reflect changes in consumers' behavior or weather conditions. Typically, daily load profiles are classified into week days and weekend days, Figure 2. Some authors consider separate analysis for each weekend day, while others deal with separate analysis for 3 types of week days, Monday, Tuesday to Thursday and Friday. In the second case the shapes of the load profiles are similar for all week days except the morning of Monday and the evening of Friday. A special type of day is the holiday. Some authors group the holidays with the weekend days.

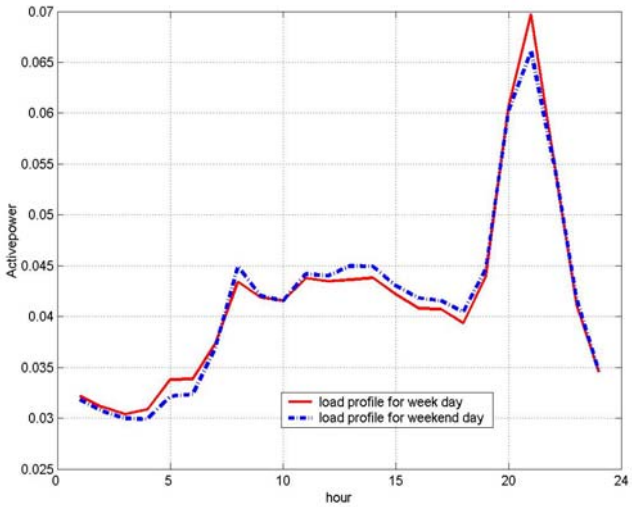


Figure 2. Load profile for week and weekend day.

The power consumption profiles of various customer types can be integrated to find the system peak loading.

The loading curve representation by 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 normal distribution law. The calculus expressions for the two characteristic sizes: mean (2) and standard deviation (3).

$$\overline{P} = \frac{\sum_{i=1}^{n} P_i}{n}, \overline{Q} = \frac{\sum_{i=1}^{n} Q_i}{n} \quad (2)$$

$$\sigma_P = \sqrt{\frac{\sum_{i=1}^{n} P_i^2}{n} - \overline{P}^2}, \sigma_Q = \sqrt{\frac{\sum_{i=1}^{n} Q_i^2}{n} - \overline{Q}^2} \quad (3)$$

where:

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

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

III. CLUSTERING METHODS

Cluster analysis is a term used to describe a family of statistical procedures specifically designed to discover classifications within complex data sets. The objective of cluster analysis is to group objects into clusters so that objects within one cluster share more in common with one another than they do with the objects of other clusters. Thus, the purpose of the analysis is to arrange objects into relatively homogeneous groups based on multivariate observations.

Although investigators in the social and behavioral sciences are often interested in clustering people, clustering nonhuman objects is common in other disciplines. For examples, clustering algorithms can be applied in many fields:

- Marketing: finding groups of customers with similar behavior given a large database of customer data containing their properties and past buying records;
- Biology: classification of plants and animals given their features;
- Libraries: book ordering;
- Insurance: identifying groups of motor insurance policy holders with a high average claim cost; identifying frauds;
- City-planning: identifying groups of houses according to their house type, value and geographical location;
- Earthquake studies: clustering observed earthquake epicenters to identify dangerous zones;
- WWW: document classification; clustering weblog data to discover groups of similar access patterns.

In this paper, the clustering algorithm is used to determine a load profile type and to analyze the demand load in a distribution substation.

It is also important to understand the difference between clustering (unsupervised classification) and discriminate analysis (supervised classification). In supervised classification, we are provided with a collection of labeled (reclassified) patterns; the problem is to label a newly encountered, yet unlabeled, pattern. Typically, the given labeled (training) patterns are used to learn the descriptions of classes that in turn are used to label a new pattern. In the case of clustering, the problem is to group a given collection of unlabeled patterns into meaningful clusters. In a sense,

[Downloaded from www.aece.ro on Sunday, February 11, 2018 at 22:44:33 (UTC) by 125.70.148.55. Redistribution subject to AECE license or copyright.]

labels are associated with clusters also, but these category labels are data driven; that is, they are obtained solely from the data.

Most cluster analyses share a similar process. A representative sample must be identified and variables selected for use in the cluster method. Samples and variables should be carefully selected so as to be both representative and relevant to the investigator's purpose for clustering. The researcher must decide whether to standardize the data, which similarity measure to use, and which clustering algorithm to select. The final stages of cluster analysis involve interpreting and testing the resultant clusters, and replicating the cluster structure on an independent sample.

It is necessary to select a clustering procedure. While the similarity or distance measures provide an index of the similarity among objects, the cluster algorithm applies a specific criterion for grouping objects together. Although researchers often disagree on the most appropriate classification scheme for cluster procedures, cluster methods are frequently classified into the following four general categories: hierarchical methods, exclusive methods, overlapping cluster procedures and probabilistic clustering. The first two categories are 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. Hierachical clustering is appropriate for small tables, up to several hundred rows. You can choose the number of clusters you like after the tree is built. Several agglomerative techniques are single link`age 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 used 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:

$$D_{KL} = \left\| \overline{X}_K - \overline{X}_L \right\|^2 \quad (4)$$

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 cunning way because a different location provides a different result. The best choice, therefore, is to place them as far as possible 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 grouping is done. At this point we need to re-calculate k new cancroids as bar centers of the clusters resulting from the previous step. After we have these k new cancroids, 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.

IV. RESULTS AND DISCUSSION

The first stage in the development of a method that would eliminate the need for hourly metering involved the use of standard curves for the consumption profile for characteristics days and the various customer groups.

Both the classification techniques and the analysis of load were tested on a set of load data recorded for a period of 6 months on the transformers within the distribution stations to the north of Romania. The load diagram is drawn using the register of watt-hour meter. The time interval of sampling load curve data is of 15 minutes and is measured from 12 midnight until 11.45 pm the following day. Therefore, the load profile is represented by 96 load values throughout the day.

The analyses were effectuated on a period of 177 days, namely 20th February – 20th April, 22nd July – 18th September and 21st October – 19th December.

Each measurement effectuated must be processed through arrangement and normalization of them. Energy consumption has been used as a normalization factor.

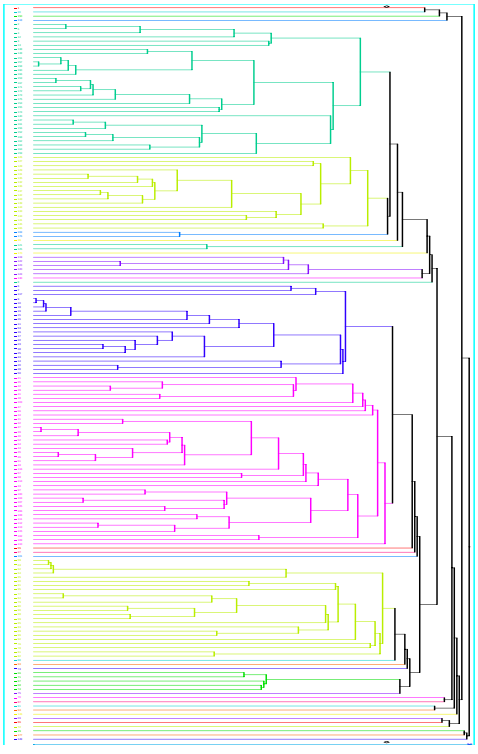


Figure 3. Dendrogram for load profile clustering.

[Downloaded from www.aece.ro on Sunday, February 11, 2018 at 22:44:33 (UTC) by 125.70.148.55. Redistribution subject to AECE license or copyright.]

This section describes the implementation of the clustering techniques. The clustered process is making progressive into coherent and representative cluster. This method was applied for the data consisting of load profile for active power during the 177 days.

Figure 3 shows the dendrogram for load profile clustering using centroid method. If we analyze this dendrogram we observe that seven clusters have resulted, which have been realized by profiles load according to seasonal time periods and to week or weekend days. Exist several days for which load profile for active power is not included in any clusters.

You can remark that a major influence on the structure of consumption is due to atmospheric conditions, especially to the environment temperature. Figure 4 illustrated the average load profile week for two periods. These are significant for demand analysis of customers and for development of strategies of planning, control a distribution networks or tariff.

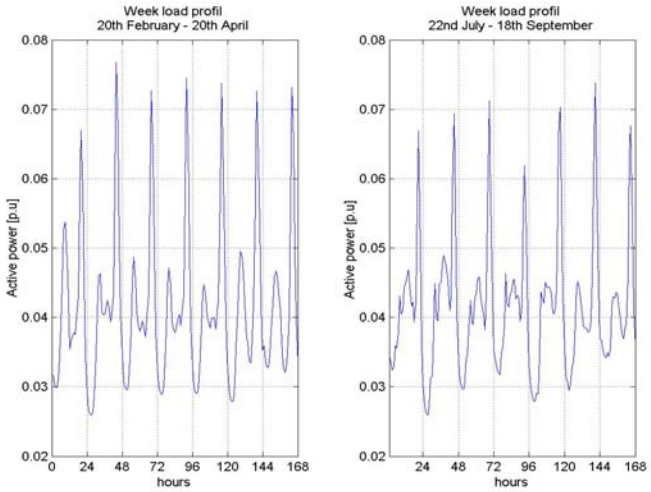


Figure 4. Average week load profile.

If the load profile is analyzed for ungrouped days, we observe that standard deviation against mean load profile is big, apart from that for one clustering day. This situation can be observed in Figure 5, which shows the active power at 3 a.m. and 8 p.m. for the days analyzed. Analyzing the temperature of load profile what remained outside groups can say that average of day is different from clustering days.

Another approach to the analysis is to consider and other factors that influenced the consumption structure, typical load profile for the consumer suppliers from this distribution substation, for example.

V. CONCLUSION

This paper describes a method for the classification of large scale sets of electric demand profiles. The load models and energy consumption of the customers served by distribution transformers are used to derive the power demand of each load bus. This method evaluates the ability of clustering classification in classifying electricity consumers based on their energy consumption.

The results obtained on several distribution stations demonstrate the effectiveness of the present method in overcoming the difficulties encountered in optimal planning and operation of distribution networks**.**

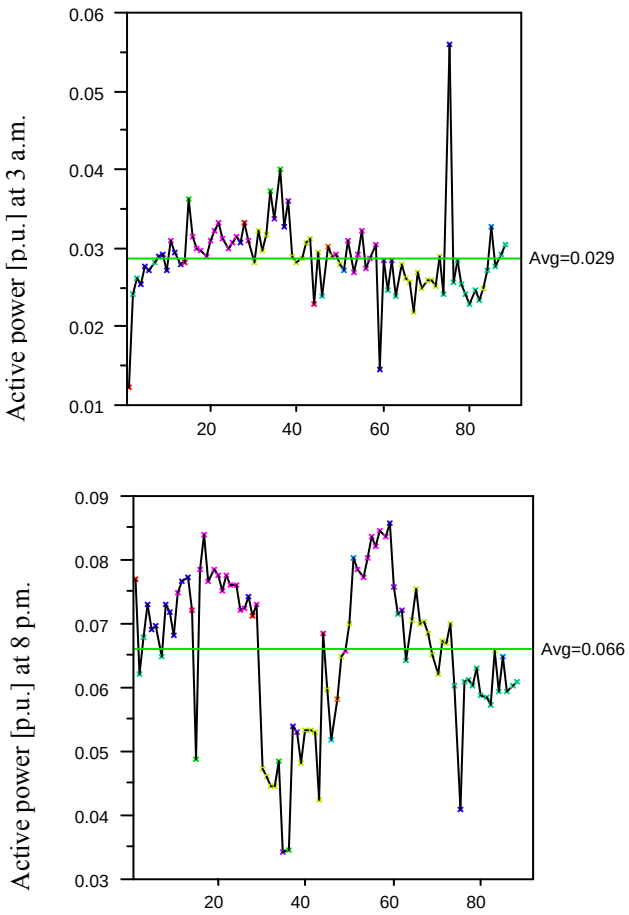


Figure 5. Active powers mean variation hourly.

REFERENCES

[1] Load profiles and their use in electricity settlement, Electricity Association, Publisher UKERC, 1997

[2] Handbook of Applied Multivariate Statistics and Mathematical Modeling, Edited by: Howard E.A. Tinsley and Steven D. Brown ISBN: 978-0-12-691360-6

[3] Gh. Cârțină, Gh. Grigoraş., E.C. Bobric, "Clustering Techniques in Load Analyse", Proc. of the International Power Systems Conference, PSC'05, 2005, Timişoara, România, pp. 123 – 130

[4] R.F. Chang, C.N. Lu, "Load profiling and its applications in power market ", Power Engineering Society General Meeting, 2003, IEEE Volume 2, 13-17 July 2003

[5] C. Nitu, A. S. Dobrescu, "The Role of Weather Indicators in Energy Consumption", Advances in Electrical and Computer Engineering, Suceava, Romania, ISSN 1582-7445, No 1/2008, volume 8 (15), pp. 17-20

[6] JMP Statistics and Graphics Guide: Version 3, SAS Institute Inc., Cary, NC, USA, 1999

[7] A.K. Jain, M.N. Murty, P.J. Flynn, Data Clustering: A Rewiew, ACM Computing Serveys, 31, 264-323, 1999

[8] Clustering: An Introduction, Available: http://www.elet.polimi.it/upload/matteucc /Clustering/tutorial_html/

[9] Gh. Cârțină, Gh. Grigoraş, E.C. Bobric, "Clustering Techniques in Fuzzy Modeling. Power Systems Applications" Casa de Editură VENUS, Iasi, 2005

[10] P.E. Sinioros, C. Filote, A. Graur, M.G.Ioannides, "A New Real Time Method of the Instantaneous Active and Reactive Power Calculus", Advances in Electrical and Computer Engineering, Suceava, Romania ISSN 1582-7445, No 1/2001, volume 1 (8), pp. 5-10

[11] M. Gavrilas, VC. Sfintes, MN. Filimon, "Identifying typical load profiles using neural-fuzzy models", 16th IEEE/PES Transmission and Distribution Conf. and Exposition, 2001, Atlanta, pp. 421-426

[12] M. Gavrilaş, Gh. Cârțină, Gh. Grigoraş, O. Ivanov. Modelarea sarcinilor din rețelele electrice, Editura PIM, Iaşi, 2006

[13] D Gerbec, S Gasperic, and F Gubina, "Comparison of Different Classification Methods for the Consumers' Load Profile Determination," presented at 17th International Conference on Electricity Distribution, CIRED, Barcelona, vol. Session 6, 2003.