BULETINUL INSTITUTULUI POLITEHNIC DIN IAŞI
Publicat de
Universitatea Tehnică „Gheorghe Asachi” din Iaşi
Tomul LIX (LXIII), Fasc. 3, 2013
Secţia
ELECTROTEHNICĂ. ENERGETICĂ. ELECTRONICĂ

CLUSTERING TO PEAK LOAD ESTIMATION IN TRANSFORMER SUBSTATIONS FROM ELECTRIC DISTRIBUTION SYSTEMS

BY

GHEORGHE GRIGORAŞ[*] and FLORINA SCARLATACHE

“Gheorghe Asachi” Technical University of Iaşi
Faculty of Electrical Engineering

Received: June 28, 2013
Accepted for publication: July 30, 2013

Abstract. The paper presents a clustering based method for consumers’ classification in representative categories characterized by the consumption information: daily energy consumption, average load and peak load. The implementation mode of method is first analysed on a working database formed by 145 small rural residential consumers from Romania. A comparative study was realized using a testing database formed by 102 consumers from same category. The obtained results demonstrated that the proposed method can be used with success in peak load estimation of transformers substations from electric distribution systems.

Key words: distribution systems; clustering techniques; peak load estimation.

1. Introduction

The estimation of loads on different parts of the distribution system is one of the most important requirements for efficient operation and planning in the power distribution systems. Estimation of loads, particularly of the peak loads, is the basis for the system state estimation and for technical and economic

[*] Corresponding author: *e-mail*: ggrigor@ee.tuiasi.ro

calculations. Load estimation, and, especially, peak load demand, influences various aspects of distribution system planning such as: transformer and conductor sizing, capacitor bank placement and so on. In order to properly analyse the performance of distribution systems, it is essential that accurate estimates of power consumption be obtained.

Data acquisition into electrical distribution systems represents a complex process, because loads are usually monitored at a few points. The main difficulties in modeling of peak loads in nodes result from the random nature of loads, diversification of load shapes on different parts of the systems, the deficiency of measured data and the fragmentary and uncertain character of information on loads and consumers. The basic constraints of the general load model in distribution networks are (Miranda *et al*., 2000; Grigoraş *et al*., 2011)

a) It must represent flows at given time, compatible with the Kirchhoff laws.

b) It must present coherency between estimated loads and measurements.

c) The load allocation must be independent with respect to the network topology under operation.

The last point is important: it would be unacceptable, from an operator point of view that the established load for a given node would "magically" chance, if he performed some switching or load transfer simulation.

In order to estimate the peak load, various mathematical formulations and methods have been developed (British, 1988; Sargent *et al.*, 1994; Broadwater *et al.*, 1997; Ionescu *et al*., 1998; Zisman *et al*., 1999; Nazarko *et al*., 1999; Cârțină *et al*., 2000; Grigoraş *et al.* 2011, 2012).

In this paper, an efficient method based on the clustering techniques is proposed. The method is based on knowledge of the consumption data about each consumer: daily energy and average power. In the first implementation step of the method a working database formed by 145 rural residential consumers is analysed. A comparative study was realized using a testing database formed by 102 consumers from same consumption category. These consumers are supplied from an electric distribution substation (20/0.4 kV) equipped with a power transformer having rated power by 63 kVA. The obtained results indicated that the proposed method leads to a good estimating of peak load in electric distribution networks.

2. Clustering Techniques

2.1. Clustering Analysis

Clustering can be considered the most important unsupervized learning problem. The scope of clustering is to determine the intrinsic grouping in a set of unlabeled data. A cluster is therefore a collection of objects which are "similar" between them and are "dissimilar" to the objects belonging to other cluster (Jain *et al*., 1999; Gasperic *et al*., 2002; Cârțină *et al*., 2005; Berkhin www). The clustering process may result in different partitioning of a data set, depending on the specific criterion used for clustering. Thus, there is a need of preprocessing before we assume a clustering task in a data set. The basic steps to develop clustering process can be summarized as follows:

a) *Feature selection*. The goal is to select properly the features on which clustering is to be performed.

b) *Clustering algorithm*. An algorithm that allows definition of a good clustering scheme for a data set must be chosen.

c) *Validation of the results*. The correctness of clustering algorithm results is verified using appropriate criteria and techniques. Since clustering algorithms define clusters that are not known *a priori*, irrespective of the clustering methods, the final partition of data requires some kind of evaluation in most applications.

d) *Interpretation of the results*. The experts compare the results of the clustering process with other data in order to draw the right conclusion.

2.2. Clustering Methods

There are two major methods of clustering: hierarchical clustering and *k*-means clustering, (Cârțină *et al*., 2005; Grigoraş *et al*., 2011; Yatskiv *et al*., 2004; Jain *et al*., 1999; Berkhin www). Hierarchical clustering is subdivided into:

a) *agglomerative* methods, which proceed by series of fusions of the *n* objects into groups;

b) *divisive* methods, which separate *n* objects, successively, into finer groupings.

Hierarchical clustering is appropriate for small tables, up to several hundred rows. Differences between methods arise because of the different ways of defining distance (or similarity) between clusters. Several agglomerative techniques will now be described in the following.

1º *Single linkage clustering (connectedness or minimum method).* The defining feature of the method is the distance between groups

$$D(r,s)=\min\left\{d\left(i,j\right)\right\}, \tag{1}$$

where object, *i*, is in cluster, *r*, and object, *j*, is in cluster, *s*.

2º *Complete linkage clustering (diameter or maximum method).* The distance between groups is now defined as the distance between the most distant pair of objects, one from each group

$$D(r,s)=\max\left\{d\left(i,j\right)\right\}, \tag{2}$$

where object, *i*, is in cluster, *r*, and object, *j*, is in cluster, *s*.

3º *Average linkage clustering.* The distance between two clusters is defined as the average of distances between all pairs of objects, where each pair is made up of one object from each group

$$D(r,s) = \frac{T_{rs}}{N_r N_s}, \tag{3}$$

where: T_{rs} is the sum of all pair wise distances between cluster, r, and cluster, s, and N_r, N_s – the sizes of the clusters r and s, respectively.

4º *Centroid Method.* In the centroid method the distance between two clusters is defined as the squared Euclidean distance between their means

$$D_{KL} = \left\| \bar{X}_K - \bar{X}_L \right\|^2. \tag{4}$$

This method is more robust to outliers than most other hierarchical methods. At each stage of hierarchical clustering, the clusters r and s, for which $D(r, s)$ is the minimum, are merged.

5. *Ward Method.* In the Ward method the Anova distance between two clusters is the sum of squares between the two clusters added up over all the variables. At each generation, the within-cluster sum of squares is minimized over all partitions which may be obtained by merging two clusters from the previous generation.

3. Determination of Peak Load of Consumers Using Clustering Techniques

An approach to peak load determination of the consumers from the LV (Low Voltage) distribution networks is presented. For this purpose the use of a hierarchic clustering method is applied to classify consumers into coherent consumption classes.

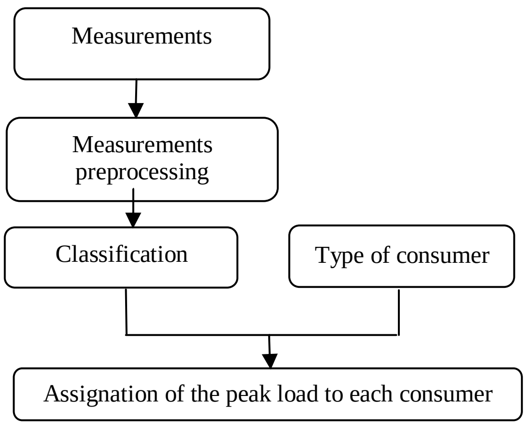

Fig. 1 – Diagram of the peak load estimation for the consumers from LV distribution networks.

The peak load of consumers is influenced by the type of consumer (residential, commercial, industrial, etc.) and, on the other hand, by the type of day or season of the year. Because a large number of consumers from various consumers' types create unnecessary problems in handling them, they could be grouped into coherent classes, seeing that some similarities exist between consumption data of these (energy consumption, average power, etc.) For each coherent consumption class, an average value of peak load is determined. The diagram of the peak load estimation for the consumers from LV distribution networks is presented in Fig. 1.

4. Peak Load Estimation of an Electric Substation

A traditional way to perform load estimation of power transformers (MV/LV) from nodes of the distribution systems is by load simulation. The method is based on the following hypotheses presented by British, (1988), Zisman *et al.*, (1999):

a) The mean loads corresponding to a cluster of nodes from the distribution system in any hour during the analysed period, is approximately proportional to the energy consumption of those nodes.

b) The loads for any hour of duration of the analysed period have a statistical distribution that can be regarded as normal.

Using these two facts, the peak load estimation of each electric distribution substation is given by the following formula (Grigoraş *et al.*, 2012):

$$P_{\mathrm{PL}} = \sum_{i=1}^{N_C} n_i \overline{P}_{\max i} + \sqrt{\sum_{i=1}^{N_C} n_i \left(\overline{P}_{\max i}\right)^2}, \ [\mathrm{kW}], \tag{5}$$

where: P_{PL} is the peak load of the electric distribution substation, [kW]; n_i – the number of the consumers from cluster i; $\overline{P}_{\max i}$ – the average maximum load of the consumers from the cluster i, [kW.h]; N_C – the number of clusters corresponding to the consumers that are supplied from electric distribution substation.

5. Case Study

To show the capability of the proposed method a database described by 145 residential consumers, belonging to a LV rural distribution network, from Romania, has been considered. The measurements of consumers' load curves were performed using electronic equipment. A sensor and an electronic device for pulse counting and data storage compose this equipment. Every measurement is a curve of 24 hourly points describing the behavior of a consumer during a day.

For each consumer were determined the information about consumption: the daily energy consumption, average load, and peak load. The consumers with missing values, outliers or energy consumption equal with zero were excluded. For clustering process, only 137 consumers were eligible. The spatial representation of these consumers is given in Fig. 2.

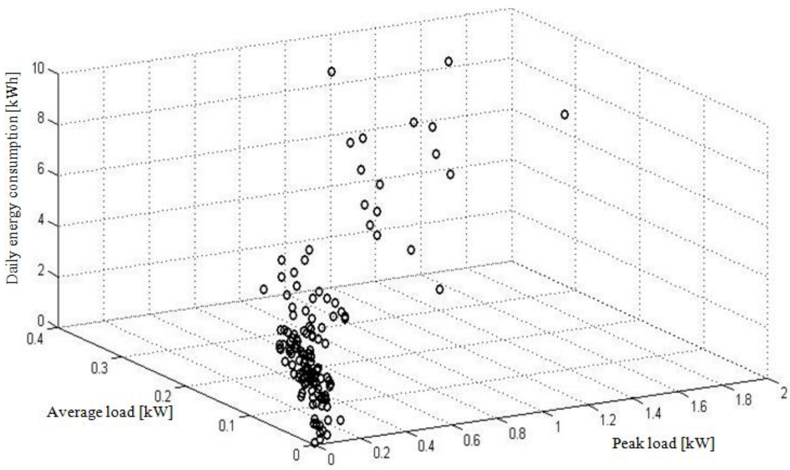

Fig. 2 – The spatial representation of the consumers in function of daily energy consumption, average load, and peak load.

In the next step, using the Ward method from the hierarchical clustering methods, the consumers were grouped in representative clusters. The consumers' characteristics from each cluster are presented in Table 1. The values of these characteristics are the medium values obtained through averaging inside each cluster.

Table 1

Characteristics of the Clusters Corresponding to the Consumers' Categories

Cluster	Number of consumers	Peak load, [kW]		Average load, [kW]		Daily energy consumption kW.h	
		Average value	Standard deviation	Average value	Standard deviation	Average value	Standard deviation
C1	19	0.17	0.034	0.07	0.003	1.58	0.066
C2	18	0.18	0.039	0.08	0.006	1.92	0.147
C3	19	0.25	0.060	0.11	0.005	2.68	0.124
C4	14	0.20	0.050	0.10	0.004	2.34	0.089
C5	18	0.15	0.040	0.05	0.006	1.14	0.156
C6	12	0.07	0.042	0.02	0.008	0.37	0.186
C7	23	0.47	0.191	0.15	0.025	3.70	0.601
C8	14	1.19	0.311	0.28	0.048	6.80	1.142

In the second step of the study, the real and estimated values of the peak load corresponding to the LV rural distribution network were determined.

Further, the method is tested on a database formed by 102 consumers from same consumption category. These consumers are supplied from an electric distribution substation (20/0.4 kV) equipped with a power transformer having rated power of 63 kVA. Each consumer was assigned to a cluster, depending on the load characteristics of this (daily energy consumption, average load) (Fig. 3).

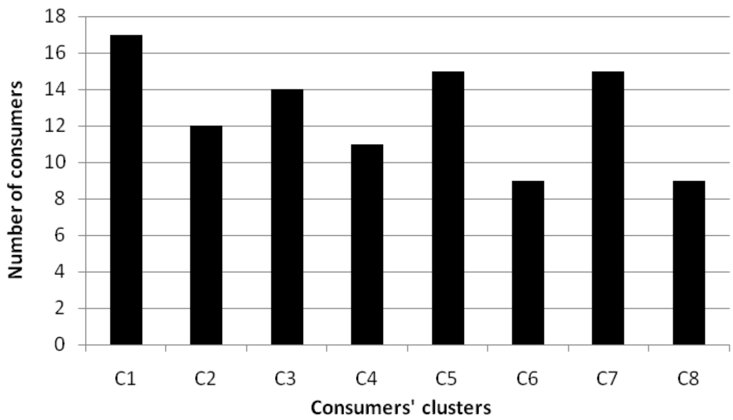

Fig. 3 – Assignation of the consumers from testing database to clusters.

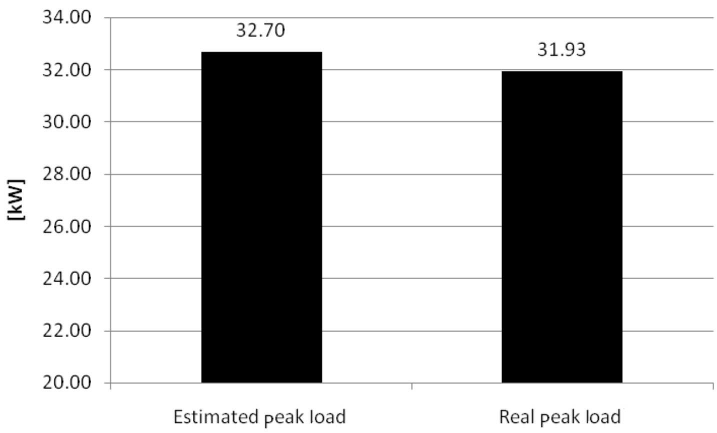

Fig. 4 – The real and estimated peak load of the analysed electric distribution substation.

Then, using information from Table 1 and relation (5), the peak load of the analysed distribution substation was obtained. In Fig. 4, the real and estimated values of the peak load obtained in the case study are presented. It can observe that the estimation error is of 2.4%.

6. Conclusions

The estimation of the loads on different parts of the distribution system is one of the most important requirements for efficient operation of electric distribution systems. In this paper, an improved approach for the peak load estimation of the electric substations, based on the clustering techniques, was proposed. The clustering techniques were applied for consumers' classification in consumption categories in function by daily energy and average from analysed distribution system are obtained using a hierarchical clustering method (Ward method). A comparison of the obtained results with the real registered data indicates that the average estimation error is of 2.4%.

REFERENCES

Berkhin P., *Survey of Clustering Data Mining Techniques.* Technical Report, Accrue Software. San Jose, California, [Online]. Available: http://citeseer.nj.nec.com/berkhin02survey.html.

Broadwater R. P., Sargent Al., Yarali A., Shaalan H., Nazarko J., *Estimating Substation Peaks from Load Research Data*. IEEE Trans. on Power Delivery, **12**, *1*, 451-456 (1997).

Cârţină G., Grigoraş G., Bobric E.C., *Tehnici de clustering in modelarea fuzzy. Aplicații în electroenergetică*. Casa de Editura Venus, Iaşi, 2005.

Gasperic S., Gerbek D., Gubina F., *Determination of the Consumers' Load Profiles.* Available: www.telmark.org/2002Sep/2-5_Gasperic.pdf.

Grigoraş G., Cârţină G., *The Impact of the Fuzzy Modeling in Electric Distribution Systems*, Lambert Acad. Publ., Saarbrücken, Deutschland, 2011.

Grigoraş G., Scarlatache F., Cârţină G., *Load Estimation for Distribution Systems Using Clustering Techniques*. Proc. of 13th Internat. Conf. on Optim. of Electr. a. Electron. Equip. (OPTIM), Transilvania Univ., Braşov, Romania, May 24-26, 2012.

Ionescu T.G., Pop O., *Ingineria sistemelor de distribuție a energiei electrice*. Edit. Tehnică, Bucureşti, 1998.

Jain A.K., Murty M.N., Flynn P.J., *Data Clustering: A Review.* ACM Computing Surveys, **31**, *3*, 264-323 (1999).

Miranda V., Pereira J., Saraiava J., *Load Allocation in DMS with a Fuzzy State Estimator*. IEEE Trans. on Power Syst., **15**, *2*, 329-534 (2000).

Nazarko J., Zalewski W., *The Fuzzy Regression Approach to Peak Load Estimation in Power Distribution Systems*. IEEE Trans. on Power Syst., **14**, *3*, 809-814 (1999).

Sargent Al., Broadwater R. P., Thompson J. C., Nazarko J., *Estimation of Diversity and KWHR-to-Peak-KW Factors from Load Research Data*. IEEE Trans. on Power Syst., **9**, *3*, 1450-1456 (1994).

Yatskiv I., Gusarova L., *The Methods of Cluster Analysis Results Validation.* Proc. of Internat. Conf. RelStat'04, 2004, 75-80.

Zisman Z., Cârţină Gh., *Application of Fuzzy Logic for Distribution System Estimation.* Proc. of the 22nd Seminar on Fundam. of Electrot. a. Circ. Theory, SPETO'99, 1999, 353-356.

* * * *Report on the Computer Program DEBUTE for the Design of LV Radial Networks.* Part **1** - *General Considerations,* Part **2** – *Program User Guide*. British Electricity Boards, Report No. *115*, 1988.

ESTIMAREA VÂRFULUI DE SARCINĂ ÎN POSTURILE DE TRANSFORMARE DIN REȚELELE ELECTRICE DE DISTRIBUȚIE FOLOSIND TEHNICI DE CLUSTERING

(Rezumat)

Se prezintă o metodă de clasificare a consumatorilor în categorii de consum caracterizate de: consumul zilnic de energie, valoarea medie şi de vârf a sarcinii absorbite, folosind tehnici de clustering. Implementarea metodei s-a făcut, într-o primă etapă, utilizând o bază de lucru formată din 145 consumatori casnici din mediul rural din România. În etapa a doua, pentru testarea metodei s-a folosit o bază test formată din 102 consumatori din aceeaşi categorie. Rezultatele obținute demonstrează faptul că metoda propusă poate fi folosită cu succes la estimarea vârfului de sarcină în posturile de transformare din sistemele electrice de distribuție a energiei electrice.