

CLUSTERING BASED APPROACH FOR CUSTOMERS' CLASSIFICATION FROM ELECTRICAL DISTRIBUTION SYSTEMS

Gheorghe GRIGORAȘ¹, Elena-Crenguța BOBRIC²

The paper presents a clustering based approach for consumers' classification in representative categories characterized by typical load profiles. For determination of the consumption categories, every customer must be characterized by the following primary information: daily (monthly) energy consumption, average load and peak load. The used database contains the daily load curves corresponding to the small customers from a rural distribution system from Romania. The obtained results it demonstrated that the proposed approach can be used with the success in building specific tariff structures of the customers or in the optimal operation and planning of distribution systems.

Keywords: clustering, customers' categories, electrical distribution system

1. Introduction

The electric load in distribution system varies with time and place. Therefore electric companies need by the accurate load data of the supply customers for distribution network planning and operation, load management, customer service and billing [1], [2]. There are several factors that influencing the customer's load:

- customer factor: type of consumption, electric heating, size of building etc;
- time factor: time of day, day of week, time of year;
- climate factor: temperature, humidity etc;
- other electric loads correlated to the target load;
- previous load values and load curve patterns.

For an electric customer, the behaviour is represented by a load profile corresponding to the electric power consumption for every period of time. Availability of such data depends on the type of customer. Generally, the small customers (like residential ones) are poorly described since a communicating meter is too expensive regarding to their consumption: for these customers there are only a few points of the curve every year. For larger customers, a

¹ Senior Lecturer, Electrical Engineering Faculty, The "Gh. Asachi" Technical University of Iasi, Romania,

² Assoc. Prof., Electrical Engineering Faculty, The "Stefan cel Mare" University of Suceava, Romania, e-mail: crengutab@eed.usv.ro

communicating meter is often available for many reasons: the billing is done every month, the consumption is high and justifies the communicating meter investment, a detailed record of consumption is necessary because prices depend on the period. Analyzed load curves may correspond to individual customer curves or to aggregates over an electric substation [2], [3].

Distribution system planners traditionally approach load modelling by estimating maximum demand values in conjunction with coincidence factors at various network levels. Although this approach has been adequate, it has several shortfalls [2], [4], [5].

The typical load behaviour during off-peak periods is unknown.

- There are inherent inaccuracies due to the use of diversity and coincidence factors which must be guessed or linked to billing information, if available.
- Energy calculations, especially losses, are not very accurate.
- Voltages at various network positions are unknown.
- Load profile dependent initiatives such as Demand Side Management (DSM) cannot be accurately modelled and evaluated.

Solving of these problems can be more efficiently using load profiles associated nodes from electric distribution networks. The power required by consumers is represented by time series, called load profile. Today, the concept of "load profile" acquires a new dimension, due to modern ways of approach and broad prospects for effective use. Load profiles are here assumed to be described by the same number of points (24 points for a daily hourly measured curve). Each profile is thus considered as a point in a p-dimensional Euclidian space (324 for daily curves). The Euclidian distance is used as a measure of dissimilarity between profiles. If this distance is not accurate, a preliminary process normalizes the profiles so that the Euclidian distance becomes appropriate [2].

In the following, it presents an approach for determination of the customer's consumption categories and typical load profiles of the consumers from these consumption categories, based on the clustering techniques. For determination of the consumption categories, every consumer must be characterized by the following primary information: daily (monthly) energy consumption, average load and peak load.

2. Clustering Techniques

Clustering techniques is the name of a group of multivariate techniques whose primary purpose is to identify similar entities from the characteristics they possess. The essence of clustering approaches is classification according to natural relationships [5], [6].

Clustering methods have as input a set of elements. An element (or feature vector, observation, or datum) is a single data item used by the clustering algorithm. These methods build a tree which approximates the similarities between elements. These similarities may be given in two ways:

- directly as a similarity or distance matrix between elements;
- indirectly: elements are described by some characteristics (attributes) and similarity between two elements is defined according to the similarities between the element characteristics.

The output of these methods is a tree where each leaf is an element and intermediate nodes represent groups of elements.

The main task of the user when performing a clustering of elements is to build a partition of the set of elements into some disjointed classes. This aspect can be illustrated by graphic presentation of a bivariate example, Figs. 1 and 2. Each point presents one element that has one value for characteristic X and one value for characteristic Y.

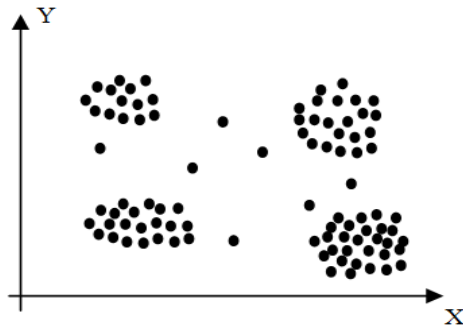


Fig. 1. Ungrouped elements

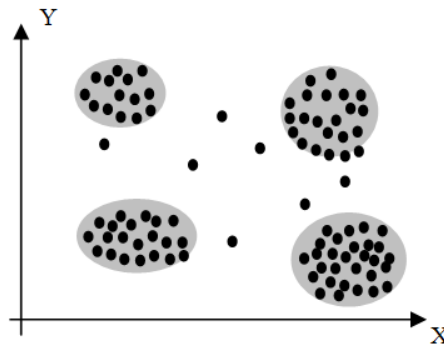


Fig. 2. Grouped elements

For instance, if the set of elements is a set of customers, classes may represent types of customers and aggregate may be the typical load profile for each class of customers.

There are some clustering methods which directly build a partition of elements (K-means method). K-means clustering is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The main idea is to define k centroids, one for each cluster so minimize an objective function. The function is a squared error function. In this case the user has to choose in advance the number of cluster and different trials are necessary to find the right number of cluster.

Other clustering methods (hierarchical methods) provide different number of clusters depending on the user. So, the result of the hierarchical clustering interpretation can be defined as follows:

- A partition of the elements into classes. Each element is described by a new characteristic which correspond to the class (an element belongs to a single class).
- Each class is summarized by a label given by the user after a careful interpretation of the characteristics of the elements belonging to the class.

3. Determination of the consumption categories

Because of the large number of small customers, sampling is the only possible way to collect data. The problem is how the selection and the analysis of the sample of customers should be made to finally get the most accurate load estimates for practical network calculations.

In this context, the electric companies require a set of load curves to represent all consumption classes. Deciding the optimal number of classes and the type of load curve for one class is a complicated problem. The practical criteria for determination of the consumption classes can be, [2], [5], [7], [10]:

- load curve variance in one class customers' class should be as small as possible;
- number of classes should be representative;
- classes should be easily linked with the database of electric company.

In order to form the consumption classes, the whole set of customers can be preliminarily partitioned into macro-categories defined on the basis of global criteria: residential, commercial, industrial, etc, [8] - [10]. Inside each macro-category, a refined classification can be identified by taking into account the daily energy consumption, average load, and peak load. Finally the typical load curves, corresponding to consumption categories are determined.

The proposed algorithm has the following major steps:

1. **Customers Research:** In this step a representative sample of the set of consumers is identified, the most relevant attributes to be measured, the cadence for data collection is defined. Finally the collected data is gathered in a large database.
2. **Data cleaning and Pre-processing:** In real problems, like this, involving a large number of measurements, spread over a large geographic area, collecting data during a considerable period of time different kind of problems will affect the quality of the database. The most relevant and frequent are communication problems, outages, failure of equipment and irregular atypical behavior of some consumers. The result will be a very large database with problems like noise, missing values and outliers. This data (after being cleaned, pre-processed and reduced) is used to obtain the division of the initial data set in classes.

3. **Partitioned into macro-categories:** The whole customer database is preliminary partitioned into consumption categories defined by the activity type: residential, commercial and industrial.
4. **Classification:** Inside each consumption category a classification into more classes, in function by the daily energy consumption, average load and peak load of the consumers is done. For realization of this classification, the Ward method is used. For every consumer from the class is determined the normalized load profile using a suitable normalizing factor (energy consumption by the period). The typical load profile for each class is obtained by averaging the values for each hour.
5. **Assignment:** Finally, to the each customer class is attributed a typical load profile according to their consumption category.

4. Determination of the consumption categories

In our study, we have considered a database described by 87 load curves corresponding to the consumers from a rural distribution network from Romania. The analyzed database belongs the residential consumer category. The measurements of individual 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 consumers were determined 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 74 consumers were eligible. Then, using the Ward method from the hierarchical clustering methods, the customers were grouped in representative clusters. The clustering process is represented by dendrogram from Fig.3.

The clusters are represented in Fig. 4 and their characteristics are presented 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 | No. consumers | Peak load [kW] | Average load [kW] | Daily energy consumption [kWh] |
|---------|---------------|----------------|-------------------|--------------------------------|
| C1 | 20 | 0.233 | 0.060 | 1.35 |
| C2 | 15 | 0.313 | 0.089 | 2.05 |
| C3 | 14 | 0.072 | 0.011 | 0.25 |
| C4 | 21 | 0.605 | 0.138 | 3.08 |
| C5 | 4 | 1.273 | 0.262 | 5.87 |

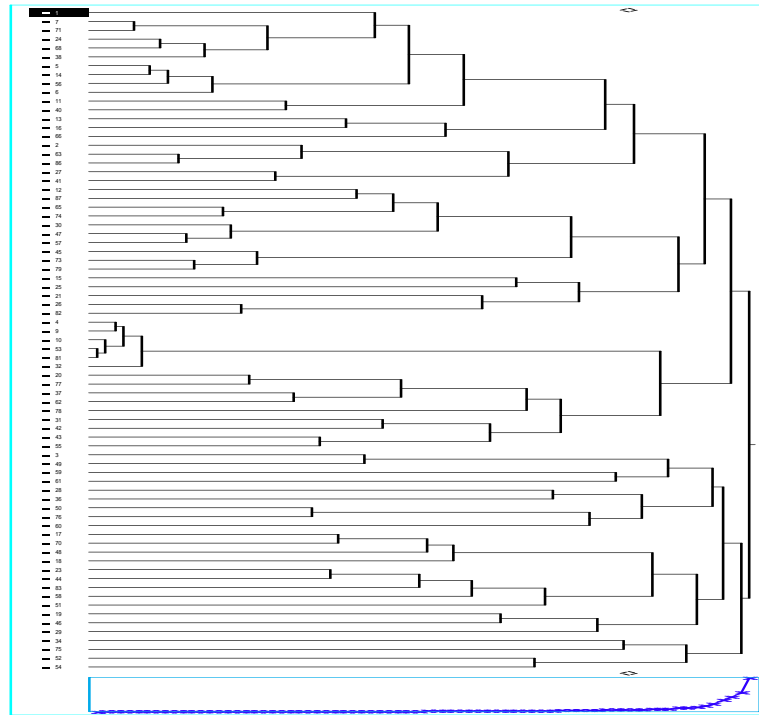


Fig. 3. Dendrogram obtained from clustering process using Ward method

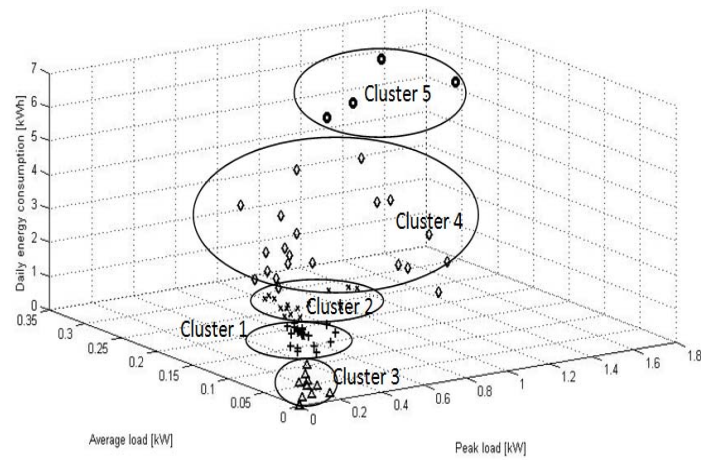


Fig. 4. The representation of the clusters obtained from clustering process

In the next step the load profiles corresponding to the consumers from each category were normalized relatively to the daily energy consumption. Then,

the average values corresponding for each consumer class were calculated.

These values transform the energy consumed by the medium member (customer) of the class in average active power demanded by it. The typical load profile for each class from the consumers' categories is obtained by representation of these coefficients, Figs. 5 – 9.

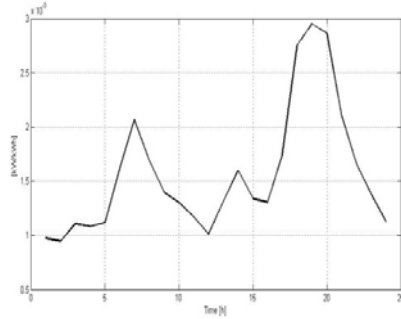


Fig. 5. Typical load profile for cluster C1

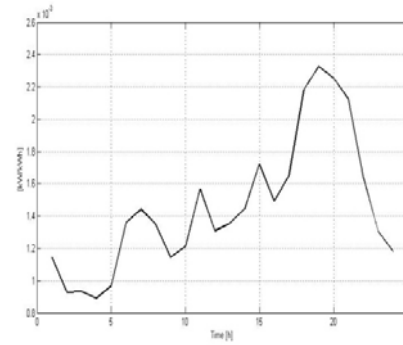


Fig. 6. Typical load profile for cluster C2

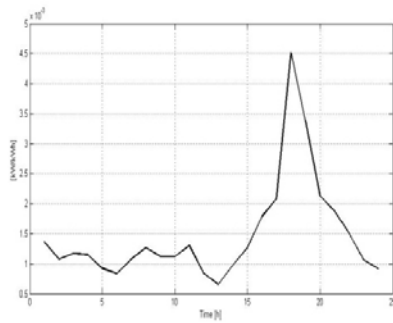


Fig. 7. Typical load profile for cluster C3

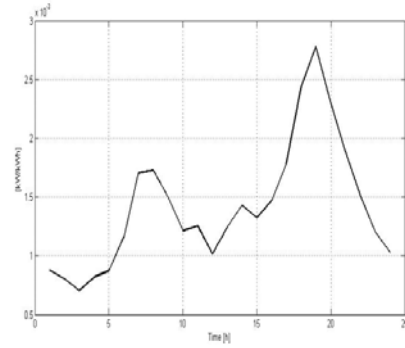


Fig. 8. Typical load profile for cluster C4

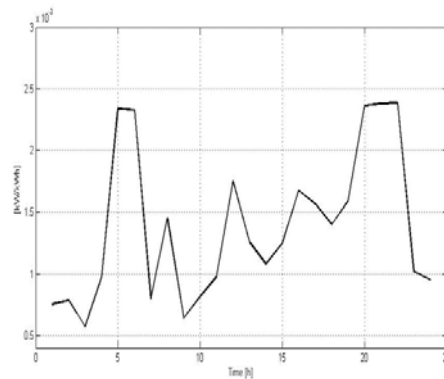


Fig. 9. Typical load profile for cluster C5

From the comparison of the typical load profiles, it results that a classification of the customer's categories is useful in view of the building of the tariff structures, demand-side management, optimal operation and planning of distribution system and so on.

5. Conclusions

In this paper a new approach, based on the clustering techniques, is proposed for determination of the consumption categories of the customers. Clustering techniques are extremely useful for assisting the distribution services providers in the process of electric customer classification on the basis of electrical characteristics (daily/monthly energy consumption, peak load, average load). The obtained results on a database of rural customers demonstrated that the methodology can be used with the success in building of the tariff structures for the customers or in the optimal operation and planning of distribution system.

REFERENCES

- [1]. *A. Seppälä*, "Load research and load estimation in electricity distribution", PhD Thesis, Helsinki University of Technology, 2006.
- [2]. *G. Grigoraș, G. Cârțină*, "The Impact of the Fuzzy Modeling in Electric Distribution Systems", Lambert Academic Publishing, Saarbrücken, Deutschland, 2011.
- [3]. *G. Hebrail*, "Practical Data Mining in Large Utility Company", available online: www.imfres.enst.fr/~hebrail/publications/hdr/Compstat_200.pdf.
- [4]. *C.G. Carter-Brown*, "Load Profile Modeling for Integrated Energy Planning", in Domestic Use of Electrical Energy Conference, 1999, pp.13 - 18.
- [5]. *G. Cârțină, G. Grigoraș, E.C. Bobric*, Tehnici de clustering în modelarea fuzzy. Aplicații în electroenergetică (Clustering Techniques in Fuzzy Modeling. Applications in Power Systems), Casa de Editură VENUS, Iași, 2005.
- [6]. *A.K. Jain, M.N. Murty, P.J. Flynn*, "Data Clustering: A Review", available online: <http://cermics.enpc.fr/~keriven/vision/articles>.
- [7]. *G. Cârțină, G. Grigoraș, C. Lupascu*, "Modeling of the customers' consumption categories in electric distribution systems", in 4th International Conference on Electrical and Power Engineering, Iasi, Romania, 2006, pp. 1175 – 1180.
- [8]. *G. Grigoraș, G. Cârțină*, "Improved Fuzzy Load Models by Clustering Techniques in Distribution Network Control", in International Journal on Electrical Engineering and Informatics, **vol. 3**, no. 2, 2011, pp. 207 - 216.
- [9]. *G. Chicco, R. Napoli, F. Piglion*, "Comparison among Clustering Techniques for Electricity Customer Classification", in IEEE Trans. On Power Systems, **vol. 21**, no. 2, 2006, pp. 933-940.
- [10]. *A. Jardini, C. Tahan, M.R. Gouvea, S.U. Ahn, F.M. Figueiredo*, "Daily Load Profiles for Residential, Commercial and Industrial Low Voltage Consumers", in IEEE Transactions on Power Delivery, **vol. 15**, no. 1, Jan. 2000, pp. 375 – 380.