ATINER's Conference Paper Proceedings Series
ENG2018-0098
Athens, 24 September 2018

Clustering Techniques for Comparative Analysis of Load Factor Profiles

Crenguta Bobric and Stefan-Gheorghe Pentiuc

Athens Institute for Education and Research
8 Valaoritou Street, Kolonaki, 10683 Athens, Greece

ATINER's Conference Paper Proceedings Series
ENG2018-0098
Athens, 24 September 2018
ISSN: 2529-167X

Crenguta Bobric, Associate Professor, University Stefan cel Mare Suceava, Romania
Stefan-Gheorghe Pentiuc, Professor, University Stefan cel Mare Suceava, Romania

Clustering Techniques for Comparative Analysis of Load Factor Profiles

ABSTRACT

The paper addresses the improving of the comparative analysis techniques of the load profiles in the electrical substations. Among the advantages of using the load profiles, it is worth noting that the simultaneity coefficients should not be estimated or calculated because the load profiles include this information and the calculation of the power losses is possible for any period of time. The analysis of load profiles can lead to credible predictions. Analysis methods based on the grouping of the load curves and their visual comparison are subjective and difficult to apply. It is proposed a method of comparative analysis of the load profiles of a distribution station, in which grouping is made based on hierarchical clustering techniques Groups analysis should also be correlated with the temperature values that can greatly influence the demand for energy. The paper presents a case study based on data collected at a distribution substation for three months, February-April. The method of spatial grouping techniques has led to a progressive grouping in coherent and representative groups, the effect of overloading or increasing the burden being modeled more precisely than in the traditional approach. Also, the transformer plots can be optimized for both peak and off-peak periods. The obtained results validate the ability of grouping techniques in the classification of the load curves and the comparative analysis of the power consumption.

Keywords: Energy consumption analysis, Load factor profile, Hierarchical clustering.

Introduction

By using modern methods of load analysis it is possible to perform load forecasts and calculate power and energy losses for variable time periods based on load profiles.

The load profiles are achieved by recording the power consumption of a consumer or a group of consumers within a time frame. The sampling period is usually hourly. In this paper, the daily load profile is represented by 24 numeric values, each value representing the consumed energy measured in one hour. Among the advantages of using load profiles we mention:

- it would not estimate or calculate the coefficients of simultaneity because load profiles include this information;
- energy and power losses can be calculated correctly at any point in the network;
- network loading and voltages are known for any time;
- transformer tap settings can be optimized for both peak and off peak periods;
- the effect of overloading or load increasing are modeled more accurately than in the traditional approach.

The load profiles are automatically made based on data periodically acquired by measuring devices connected to computers that continuously monitor the state of the energy network. One problem that exists is how to process these time series of values so that the high level information can be extracted in order to make it possible to obtain the above mentioned benefits.

One solution is represented by using methods and techniques from pattern recognition and machine learning to identify similar behaviors and to use the obtained information in other higher level processing.

Pattern Recognition in Load Analysis

Pattern recognition is a field with a spectacular evolution that seeks to identify similar relationships between the abstract representations of entities or phenomena from the real world. In this paper we will consider as in the book [1] a pattern being a representation based on the results of the measurements and / or observations made on the object or phenomenon in the real world.

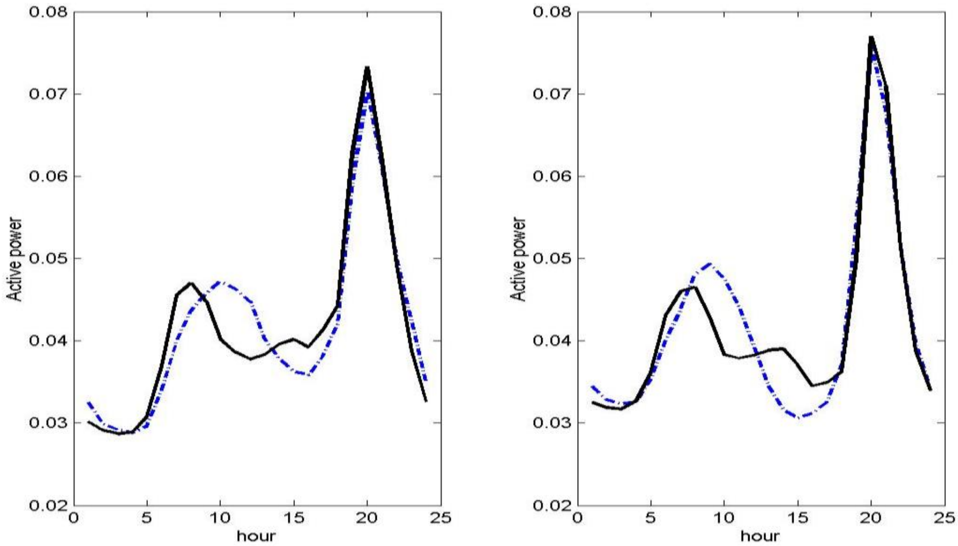
$$\mathbf{x} = (x_1, x_2, \ldots, x_p) \tag{1}$$

where the x_i components of vector $\mathbf{x}$ are called features.

The set **E** of patterns is a finite set. In the model we approach, it is considered that the similar patterns are grouped in subsets that form a $P_M(E)$ partition in equivalency classes of **E**. This model has two restrictions: all forms are classifiable, and a form can only be part of one class. The advantages of this model in the case of the load analysis overcome the inconveniences induced by these restrictions.

In our paper pattern will be considered as being an array with the 24 values of the hourly consumed energy over a day. Each day will be represented by a pattern like this. For each consumer, each day has its specificity in terms of electricity consumption and consequently each pattern has its own shape (Figure 1).

Figure 1. *Load Profile in case of 2 Consumers [5]. Continuous Line - Working Day, Dashed Line – Weekend Day*



The set of patterns represents the set of loading profiles for the studied period (177 days). The question is to identify groups of similar patterns in our dataset, how many are they, and from which patterns (daily profiles) are composed.

Clustering

Clustering belongs to the unsupervised classification algorithms. The aim is to group the objects into clusters where each group containing objects with a degree of similarity between them closer to other items in other groups

The aim of the clustering algorithms is to estimate the density or a structure of a data set without previous information about their classification [3]. The issue of these algorithms will be a partition or a string of partition of pattern set E.

The problem of finding the optimal partition is NP complete, and various heuristic algorithms have been published.

The well-known algorithms of unsupervised classification are K-means, Dynamics Nucleus, Hierarchies, and other distance based algorithms.

By applying different algorithms on a same data set it, different partitions may be obtained, because these partitions are obtained based on different optimization criteria. The problem is how to evaluate the multiple partitions issued from various clustering algorithms or from the various running of the same algorithm with different parameters. The aim is to find an optimal and consistent partition of the pattern set.

Central Partition [2]

Let a string of partitions P^1 (M), P^2 (M), ... , P^K (M), of a pattern set E, in M classes, issued from an algorithm executed by various parameters, or issued by applying different algorithms on the same pattern set **E**.

Each of these partitions may be represented by integer arrays γ^k, where its elements, γ^k_i, represents the index of the class that belongs the pattern *i* in partition P^k(M). A partition P^i in M equivalence classes is generated by the equivalence relation:

$$u_i(x,y) = \begin{cases} 1 \; dacă \;\; \gamma_x^i = \gamma_y^i \\ 0 \quad altfel \end{cases} \qquad (2)$$

Definition. The central partition of a string P^1 (M), P^2 (M), ... , P^K(M), is the partition P^* (M) that is situated at the minimum distance from P^1(M), P^2 (M), ... , P^K (M).

Several metrics can be defined over the partition set. We will use the proposed metric in [1]:

$$\mathrm{d_C}(P^i, P^j) = \frac{1}{2} \sum_{(x,y)\in E\times E} \left| u_i(x,y) - u_j(x,y) \right| \qquad (3)$$

It is to note that the values to be summed are equal to the value of the program expression:

```
iclasMat[i][x] == iclasMat[i][y]
XOR iclasMat[j][x] == iclasMat[j][y]
```

where the operator == produces value 1 for true and 0 for false.

The problem of determining the central partition can be formulated as follows: to find the relation of equivalence u * (x, y) for which

$$\sum_{i=1}^{K} \sum_{(x,y)\in E\times E} \left| u^*(x,y) - u_i(x,y) \right| = \min \qquad (4)$$

The problem is NP difficult, you will try to apply a heuristic to find the partition that achieves the best consensus with the partition string $\boldsymbol{P}^1$, $\boldsymbol{P}^2$, ... , $\boldsymbol{P}^K$.

Definition. Let the equivalence relation:

$$w^K(x,y) = \begin{cases} 1 & dac\breve{a} \ \sum_{i=1}^{K} u_i(x,y) = K \\ 0 & altfel \end{cases} \tag{5}$$

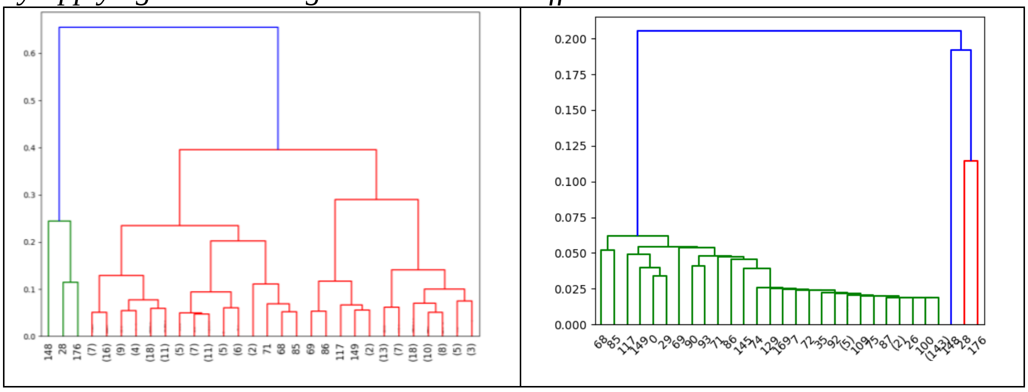
Two pattern are in relation w^K if that belong to the same class in each of partitions $P^1(M)$, $P^2(M)$, ... , $P^K(M)$.

The partition generated by the relation w^K is called the partition of the strongest patterns $\Pi_S(E) = \{\Gamma_1,..., \Gamma_S\}$, where Γ_i are the strongest patterns.

Results

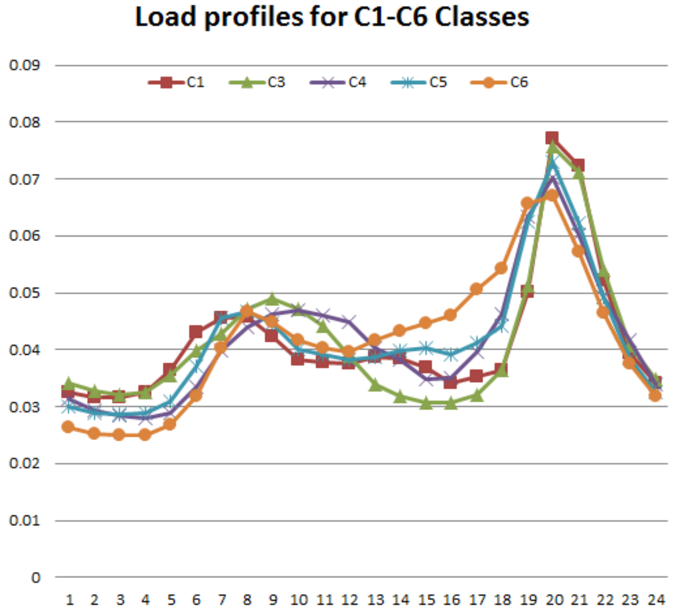
The number of patterns representing the load profile in 177 days was classified by applying an ascending hierarchical classification algorithm using different metrics and linkage techniques. It is reminded that a hierarchy over a set **E** is a set of sets **H** with the properties: (i) **E** belongs to **H**, (ii) whichever is x of **E**, the set {x} belongs to **H**, and (iii) for any two sets h, h' of **H**, if their intersection is not empty then either h is included in h ', or h' is included in h [1]. The well-known graphical representation of hierarchies is dendrogram (Figure 2).

Figure 2. *Two Dendrograms of the Same Set of Patterns (Load Factors) Obtained by Applying the Same Algorithm but with Different Parameters*



The partition string obtained by repeated application of the hierarchical agglomerative algorithm with different parameters [4], was analyzed by the central partition method, resulting the strongest patterns. The following types of load factors were determined (Figure 3).

Figure 3. *Load Profiles for Resulted Pattern Classes*



It can be noted in Figure 3 the 6 types of load factor profiles evolutions, corresponding to specific conditions of the analyzed power system.

The shapes belonging to each of the 6 classes and the days of the data were identified (in Figure 4 the calendar days are colored according to the classes of load profiles).

Figure 4. *Coloring the Days in the Calendar according to the Corresponding Load Factor Class*

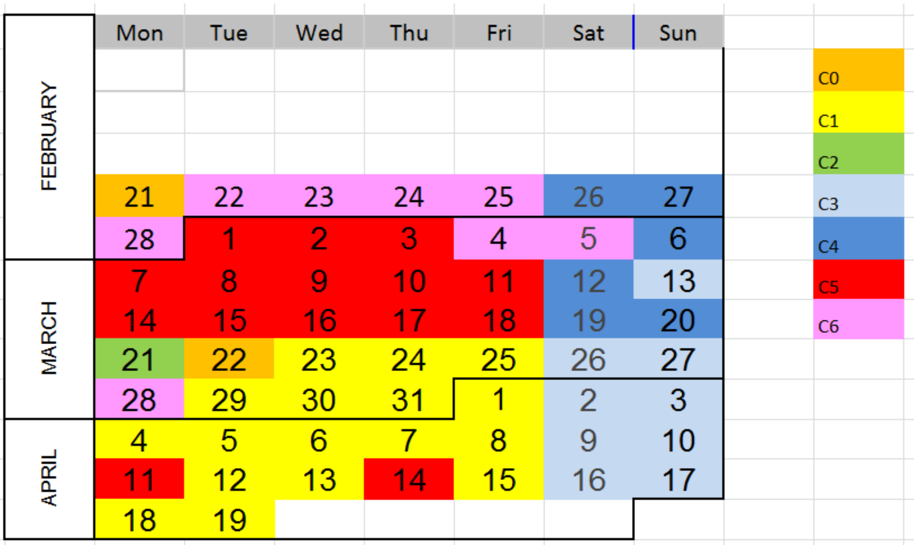


Figure 6 shows a subset of the 177 load profiles, and the C0 class corresponds to the days when, due to the intervention on the measuring and control equipment, load factors registration was incomplete. This is a supplementary verification of the execution of clustering algorithms that have been able to be isolated and typical days.

Conclusions

The paper presented a method of comparative analysis of the load profiles of a distribution station, registered in a period of two months, between 21 February - 19 April.

The grouping of load profiles by visual comparison is subjective and difficult to apply, so there were applied techniques and algorithm from unsupervised pattern classification.

The hierarchical clustering algorithms lead to a progressive grouping in coherent and representative groups. The dendograms highlight grouping on characteristic days, working days and rest days, pregnancy curves, for the period of analysis.

Temperature is a factor of influence of the consumer demand, which is evidenced by the analysis of the dendograms obtained from the grouping process.

The results demonstrate the ability of grouping techniques in the classification of load curves and the comparative analysis of consumption.

The analysis of a wider period implies a big number of data, and in consequence faster algorithms are required. Further researches will be oriented toward exploring the parallel computing potential of clustering algorithms that use the SIMD operations of the PowerXCell 8i microprocessor [6]. It is planned to develop a bi-level parallelization version for the algorithm of finding Central Partition for the cluster USV-RoadRunner (6.5 Tflops) as it is underlined in [7].

References

[1] Stefan-Gheorghe Pentiuc, Pattern Recognition. Methods, programs, applications, Ed. Universitatii Stefan cel Mare Suceava, p. 255. ISBN 973-97787-5-5.
[2] Celeux, G.; Diday, E.; Govaert, G.; Lechevalier, Y. and Ralambondrainy, H. (1989). Classification Automatique Des Donnees, Dunod, Paris.
[3] Bobric, Elena C., Gheorghe Cartina, and Gheorghe Grigoras. "Clustering techniques in load profile analysis for distribution stations." Advances in Electrical and Computer Engineering 9.1 (2009): 63-66.
[4] scikit-learn - Machine Learning in Python, http://scikit-learn.org.
[5] Cartina, G., Grigoras, G., Bobric, E. C. & Comanescu, D. (June 2009). Improved fuzzy load models by clustering techniques in optimal planning of distribution networks. In PowerTech, 2009 IEEE Bucharest (p. 1-6). IEEE.
[6] Pentiuc, Stefan-Gheorghe, and Ioan Ungurean. "Multilevel Parallelization of Unsupervised Learning Algorithms in Pattern Recognition on a Roadrunner Architecture." Intelligent Distributed Computing V. Springer Berlin Heidelberg, 2012. p. 71-80.
[7] Ungurean, Ioan, Ionela Rusu, and Stefan-Gheorghe Pentiuc. "High-performance computing on a supercomputer based on new-generation processors." Tier 2

Federation Grid, Cloud & High Performance Computing Science (RO-LCG), 2012 5th Romania. IEEE, 2012.