Electric Power Systems Research 141 (2016) 114–123

Contents lists available at ScienceDirect

Electric Power Systems Research

journal homepage: www.elsevier.com/locate/epsr

Comparison and clustering analysis of the daily electrical load in eight European countries

Pietro Ferraro*, Emanuele Crisostomi, Mauro Tucci, Marco Raugi

Department of Energy, Systems, Territory and Constructions Engineering, University of Pisa, L.go Lucio Lazzarino 1, 56122 Pisa, Italy

ARTICLE INFO

Article history:
Received 31 January 2016
Received in revised form 22 June 2016
Accepted 5 July 2016

Keywords:
Clustering
Daily load
Expectation maximization
Spectral clustering
Fuzzy c-Means

ABSTRACT

This paper illustrates and compares the ability of several clustering algorithms to correctly associate a given aggregate daily electrical load curve with its corresponding day of the week. In particular, popular clustering algorithms like the Fuzzy c-Means, Spectral Clustering and Expectation Maximization are compared, and it is shown that the best results are obtained if the daily data are compressed with respect to a single feature, namely the so-called "Morning Slope". Such a feature-based clustering appears to outperform the clustering results obtained upon using other classic features, and also with respect to using other conventional compression methods, such as the Principal Component Analysis, in all the examined European countries. This result is particularly interesting, as this feature provides a direct physical interpretation that can be used to obtain insights on the structure of the daily load profiles.

1. Introduction

The recent increase of non-dispatchable energy sources, that inject power into the grid in a non predictable way, has led to several stability issues in the electrical grid and it is one of the reasons to carefully study and analyse the electrical load consumption data: in order to maintain the balance between produced and consumed energy, several plants (e.g., thermoelectric power plants) are continuously maintained switched on at a low level, as a backup to match the energy demand, if needed (i.e., if they were switched off, then it would take an exceedingly long time before they could effectively be used as a backup). Such an operation is very expensive, especially as it might be very rare that they are used in practice. In this framework, it becomes of primary importance to be able to fully grasp the behaviour of the electrical load: to understand in a clear and quantitative way which parameters characterize the daily load profiles, thus to be able to predict, in an accurate manner, the electrical demand. This will provide useful insights to the energy suppliers in order to schedule the operation of power plants in a more efficient, and economically convenient way; moreover, the recent political scenario is shifting toward a stronger deregulation of the energy market: this will allow in a near future the single user to buy energy at a variable price, thus making even more important to have a clear understanding of the electrical load behaviour, for economical reason.

In order to fully characterize the electrical load, many mathematical methods have been employed during the past few decades (Fourier or Wavelet analysis to make an example) resulting in a very deep knowledge of the subject by the scientific community. Despite this, such analyses often employ techniques whose results due to their mathematical content, might prove difficult to be interpreted by a non technical audience; for this reason it appears useful to employ methods whose results have a clear physical meaning that make them understandable by anyone lacking depth of knowledge in the subject at hand. In this perspective, a parameter that plays a crucial role in analysing the daily electrical load is the day of the week: different behaviours can be observed in working days (e.g. Mondays to Fridays), preholidays (e.g. Saturdays and days prior to festivities) and holidays (e.g. weekends, Christmas and other festivities), while similar trends can be obtained comparing days that belong to the same set. From these considerations, clustering algorithms, due to their ability to find common patterns and due to their natural physical interpretation, appear as optimal candidates.

1.1. Clustering and classification

Classification and clustering of time series signals is an important area of research in several fields. Clustering refers to the ability to aggregate similar objects together, in groups called *clusters*. The

* Corresponding author.
E-mail address: ferr.pietro@gmail.com (P. Ferraro).

http://dx.doi.org/10.1016/j.epsr.2016.07.002
0378-7796/

idea of similarity is fundamentally, a human one: it is not trivial to define, in a rigorous way, what *similar* means. Moreover, even if a definition could be found, it would not be an unambiguous one, since it would still depend on the metric used to compare the data. There are numerous motivations to group objects into clusters, an in depth analysis can be found in [1]; in this paper the reasons to use clustering algorithms are

- they have a good predictive power;
- they allow to compress the data into a reduced number of information.

These properties lead to a more efficient description of the data, which improves the ability to choose the actions to take in specific situations.

1.2. State of the art

The use of clustering techniques to analyse the electrical load data is a consolidated area: in [2] clustering algorithms are used in order to analyse and divide large electricity customers into classes to estimate their typical days and their representative daily load profiles, in [3] clustering techniques are used to propose an annual framework for optimal price offering. For this purpose, load profiles of customers are used as well as their consumption patterns. In [4] a new clustering algorithm for load profiling is proposed, based on billing data. In [5] an analysis of the load profiles of a representative sample of Spanish residential users is performed by using dynamic clustering (i.e., dynamic in the sense that the load profiles are interpreted as a time series database). A similar attempt has been previously performed, in [6]: the paper shows how to classify electrical customers, in particular, 234 non residential customers in Italy connected to the Medium Voltage (MV) distribution system, using different clustering algorithms; the provided results are not interpreted. In [7], a building in a university campus in Greece is analysed through a clustering analysis. In some references, clustering analysis is mainly performed as a basis for load forecast. Among others, we remind Refs. [8–11]. In particular, in [12], the authors point out that clustering several consumers can lead to an increase in forecasting accuracy. Other related works include [13] in which clustering algorithms are employed to study the electrical load profile and for peak load assessment; in [14] an initial set of centroids, defined by a user defined centroid model, is used to identify load patterns; in [15] the residential electrical load is modelled using mixture model clustering and Markov models; finally in [16] the authors propose a neuro fuzzy classification methods to monitor the load in a non intrusive fashion.

1.3. Objective of the paper

In the works cited so far, and in the literature in general, very little was done to identify *aggregate* electrical daily patterns using clustering algorithms. It is clear, intuitively, that during the week different electrical behaviour can be observed and classified; the objective of this paper is to provide a mathematical framework to formally describe how the classification can be performed automatically and up to what extent.

A preliminary analysis along the previous lines is described in [17] where it is shown that the task of classifying the data in two different clusters can be performed exceedingly well, obtaining consistent performances of about 95% for each investigated country. The same algorithms though, do not provide the same outcomes for the three classes case, resulting in very poor performance for what regards a direct approach. In the aforementioned paper it was shown that the results can be improved, ranging from 89% to 94%, using a hierarchical approach; in Section 3 this aspect is further investigated.

In both cases though, these performances were not obtained for the direct raw values, for which the results were extremely poor, but on a set of features extracted from the dataset, considered one by one.

This paper extends the work done in [17] along a number of different lines:

- The comparison is extended to more countries of the European union;
- The preliminary results of [17] are rigorously confirmed by using different clustering algorithms;
- We consider PCA as an alternative well-established compression technique to pre-treat data before running a clustering algorithm, and we compare the results with those obtained in [1] using single features (see later Table 2 for a full list of single features);
- A sensitivity analysis is performed on the most informative feature, showing its reliability and robustness.

This paper shows that the feature called *Morning slope* performs better even than a well established compression method like the Principal Component Analysis (PCA), no matter what clustering algorithm is employed, with the advantage that such a feature preserves a clear physical meaning as it is focused on the load values during specific hours of the day. This result, which to the best of the authors' knowledge has never been noticed before, is rather surprising since it appears to hold in very different countries that present load data that are hardly comparable due to different sizes, latitudes and habits (in France, for example, where electrical energy is also used for heating purposes). This classification, being more accurate than the calendar (as it analyses directly the load curve), can be used as a preliminary analysis for daily load forecasting algorithms in which a different prediction model is used for each cluster [18]. Moreover, on the basis of this classification, average day profiles have been found (see Section 3 for details) which could help energy suppliers to tailor their tariffs.

1.4. Organization of the paper

This paper is organized as follows: Section 2 introduces the used database, provides some initial insight on the available data, and shortly illustrates the used clustering algorithms and the adopted performance indices. Section 3 thoroughly compares the hourly load patterns among the considered countries in different cases, Section 4 shows the results and the performances of the clustering algorithms and Section 5 provides a detailed description of the differences between the use of the calendar and our analysis. Finally, Section 6 summarizes the paper findings.

2. Background: electrical load data and clustering algorithms

2.1. Data set

The data used in this paper are taken from the electrical load data freely available from the ENTSO-E database.[1] ENTSO-E is the European Network of Transmission Systems Operators for Electricity. The ENTSO-E statistical database includes a range of historical data sets regarding power systems of ENTSO-E member Transmission Systems Operators (TSOs). Following the merging of former TSOs associations in 2009, ENTSO-E has become the single data

[1] https://www.entsoe.eu/data/data-portal/consumption/Pages/default.aspx