Sustainable Cities and Society xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Sustainable Cities and Society

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

Using clustering techniques to provide simulation scenarios for the smart grid

Pedro Miguel[a,c,*], José Gonçalves[a,c], Luís Neves[a,b,c], A.Gomes Martins[a,c]

[a] Energy for Sustainability Initiative, University of Coimbra, Coimbra, Portugal
[b] School of Technology and Management, Polytechnic Institute of Leiria, Leiria, Portugal
[c] INESCC—Institute for Systems Engineering and Computers at Coimbra, Coimbra, Portugal

ARTICLE INFO

Article history:
Received 29 September 2015
Received in revised form 16 March 2016
Accepted 18 April 2016
Available online xxx

Keywords:
Data clustering
Demand response
Energy box
Energy storage
Smart grid
Distribution system operator

ABSTRACT

The objective of this work is to obtain characteristic daily profiles of consumption, wind generation and electricity spot prices, needed to develop assessments of two different options commonly regarded under the smart grid paradigm: residential demand response, and small scale distributed electric energy storage. The approach consists of applying clustering algorithms to historical data, namely using a hierarchical method and a self-organizing neural network, in order to obtain clusters of diagrams representing characteristic daily diagrams of load, wind generation or electricity price. These diagrams are useful not only to analyze different scenarios of combined existence, but also to understand their individual relative importance. This study enabled also the identification of a probable range of variation around an average profile, by defining boundary profiles with the maximum and minimum values of any cluster prototypes.

1. Introduction

Abundant recent developments regarding power systems are focused on the concept of smart-grids (Bundesnetzagentur, 2011; Electric Power Research Institute, 2008; European Union Advisory Council, 2010; Pacific Northwest National Laboratory, 2010; SmartGrids, 2012), aiming to achieve the ultimate goal of balancing supply and demand, not only by acting on a strictly controllable supply side in response to the naturally uncontrollable demand, although predictable to a certain degree, as in the past, but now also using new solutions to make the demand respond to the less predictable variations of supply, that are characteristic of renewable sources, or increasing the ability to store energy, so that energy generated in one period of time can be used in a period of time when demand really occurs.

In this context, there is an obvious need to evaluate the true value of these options so that the required investments can be justified. But to obtain figures of the significance both options can achieve, there is an absolute need of data, namely on the time-dependent behaviour of load demand, renewable generation and electricity price.

The concept of Demand-response (Albadani e and El-Saadany, 2008; Sá, 2011; Mahmoudi, Eghbal, & Saha, 2014), is based on the possibility of inducing consumers to dynamically change their consumption pattern in order to match the pattern that can be most convenient to the grid operator. The partial accomplishment of this objective can result of the use of stimuli as, e.g., a Real Time Price (RTP) scheme (Nunes, 2011) in which the price of electricity seen by the individual consumer changes frequently. New proposed devices can use this stimuli to automatically decide on behalf of the consumer if certain loads can be temporarily stopped, or their start postponed, in order to avoid a price peak. But for the massive deployment of these devices, a careful evaluation is needed in order to justify the heavy investment in all the infrastructure.

Likewise, the search for new storage options led to the concept of distributed storage (Ribeiro, Johnson, Crow, Arsoy e, & Liu, 2001; Divya and Østergaard, 2009; Lassila, Haakana, Tikka, & Partanen, 2012; Sigrist, Lobato e, & Rouco, 2013; Leou, 2012), e.g., using the batteries installed on the electric vehicles that are expected to flood the electric grids in the near future with the so called Vehicle-to-grid (V2G) option (Lassila et al., 2012). But this, or any other distributed storage option, needs also to be assessed in order to justify a significant investment in the supporting infrastructure, and this needs to compare the temporal evolution of demand, prices and

* Corresponding author at: University of Coimbra, Energy for Sustainability Initiative, 3000-033 Coimbra, Portugal.
E-mail addresses: pcmiguel@gmail.com (P. Miguel), jagoncalves@gmail.com (J. Gonçalves), luis.neves@ipleiria.pt (L. Neves), agmartins@uc.pt (A.Gomes Martins).

http://dx.doi.org/10.1016/j.scs.2016.04.012
2210-6707/

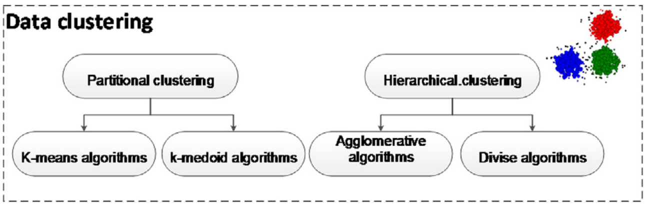

Fig. 1. Data clustering techniques (Pujari et al., 2001).

also renewable generation with time, if the aim is to help improving the match between generation availability and demand.

The development of research work (Miguel, Neves, & Martins, 2014; Gonçalves, Neves, & Martins, 2015) related with the two options above, led to the need of obtaining representative daily diagrams of demand, renewable generation (wind) and electricity gross market prices, in order to generate reliable scenarios of analysis, aiming also to obtain a measure of the representativeness of each diagram regarding the context of a whole year. To achieve that goal, different clustering techniques were applied to the historical data sets of hourly consumption, wind generation and gross market prices of a complete year. The present paper, focuses on this process of obtaining representative daily diagrams or profiles.

The next section makes a short review of clustering techniques, justifying the choice of the particular methods that were used and explaining their application to the problem. Section 3 presents the application of the chosen methods to the data used, representing the consumption of a small town in Portugal, the wind generation data provided by the Independent System Operator (ISO) and the gross market prices obtained from the wholesale electricity market operator website. Section 4 presents the results that were obtained, and finally, Section 5 presents conclusions.

2. Clustering techniques

2.1. Motivation for using clustering methods

The present trend towards the development of smart grids presents an opportunity for data clustering (Benítez, Quijano, Díez, & Delgado, 2014), considering the large data sets for the dynamic analysis of the electricity grid, such as: the annual monitoring of distribution and residential load demand, energy market prices and renewable generation. Although the evaluation of whole years is mandatory, the need to simulate management strategies requires some sort of daily management leading to the need of daily characteristic profiles. e.g., a number of extraordinary days may particularly justify the use of demand response and energy storage, namely when spot prices achieve maximum values and there is potential to shift usage, or when renewable generation is totally decoupled from consumption. But, for assessing their relevance, it is important not only to have models of the diagrams of consumption, generation and market price evolution but also to assess their relevance for a given period of time.

Clustering methods may provide a response for this need by performing an automatic classification of daily diagrams over a long period of time.

Cluster analysis is a convenient method for identifying homogenous groups of objects or clusters (Mahmoudi et al., 2014; Mooi and Sarstedt, 2011), aiming to group objects by similarity, but keeping a significant difference between groups. Clustering techniques can be divided in two main groups, Partitional clustering methods and Hierarchical clustering methods (Pujari, Rajesh, & Reddy, 2001), as presented in Fig. 1.

Partitional algorithms use an interactive optimization paradigm. They start with an initial partition and, in each iteration, objects are swapped if the quality of the clustering is improved, by applying an interactive control strategy. The definition of the clusters will strongly depend of the initially selected partition. Another relevant issue is the representation of clusters. In the K-mean algorithms a cluster is represented by its center of gravity while the K-medoid algorithms use the closest object from the center as the center representation. However, according to Sousa (2006) the limitations of this method are the dependence on the prototypes initially assigned to clusters and also the lack of a clear assignment of objects to clusters, creating difficulties to the definition of prototypes.

The Hierarchical methods (HM) create sequenced partitions, each one containing the previous one in a hierarchical way. There are two types of hierarchical methods (Maimon and Rokach, 2005), the agglomerative clustering algorithms and the divisive clustering algorithms, the former performing a bottom-up approach and the latter a top-down approach, using a similar clustering rule based on similarity or distance (Maimon and Rokach, 2005). According to Sousa (2006), the agglomerative clustering algorithms are probably the most used hierarchical methods, and therefore one of the choices made for the current work was based on such methods. A more complete description is then provided in Section 2.2.

An alternative procedure introduced by Teuvo Kohonen in the early1980's, the self-organizing maps (SOM), is a type of artificial neural network (ANN) with an unsupervised learning process and unknown output partitions (Kohonen, 1998). The Kohonen SOM could be compared to conventional clustering methods, considering the internal allocation rules and their performance (Sousa, 2006).

Other specialized procedures can be found in the literature, as the weighted fuzzy average (WFA) K-means algorithm and the "Modified Follow the Leader" (Mahmoudi-Kohan, Moghaddam, & Bidaki, 2009).

According to Mahmoudi-Kohan et al. (2009) WFA K-means is well suited to design electricity tariffs, while "Modified Follow the Leader" is suited for the selection of demand response policies.

For the current work the authors chose to implement the HM and SOM methods due to their adequacy to the objective of clustering of load diagrams, as shown by Sousa (2006).

2.2. Hierarchical method (HM)

According to Sousa (2006), the hierarchical clustering method investigates possible groupings of data by creating a structure similar to a hierarchical tree. Such tree is comprised not only by a simple set of clusters, but rather by a multiple level hierarchy, where clusters at one level are grouped in clusters of a higher level. The basic procedure for developing the cluster prototypes is performed according to the following steps (Sousa, 2006):

The input data is evaluated and a decision is made on what kind of clustering results are valuable to obtain later on. For the cases studied on this paper, this decision is based on either the amplitude difference between cluster members (defined as the difference between extreme values of a daily profile) or on the diagram shape. For the price diagrams it was assumed that what is relevant to differentiate prototypes is the amplitude difference, while for electricity profiles (consumption and generation) what is relevant is the diagram shape. As stated earlier, a possible approach of the hierarchical method consists in merging objects according to their linkage distance, using one of multiple possible definitions, namely: single, complete, average, centroid or ward distance, the latter being used in the present case. After calculating distances between data members, it is possible to group objects according to their proximity (the shortest distance). As new higher level clusters are created, comprised by more members, it is necessary to determine the distances between them until the hierarchical tree is composed by only one

Please cite this article in press as: Miguel, P., et al. Using clustering techniques to provide simulation scenarios for the smart grid. *Sustainable Cities and Society* (2016), http://dx.doi.org/10.1016/j.scs.2016.04.012