Expert Systems With Applications 86 (2017) 54–63

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Electricity clustering framework for automatic classification of customer loads

Félix Biscarri*, Iñigo Monedero, Antonio García, Juan Ignacio Guerrero, Carlos León

Escuela Politécnica Superior, Electronic Department, Virgen de Africa, 7, Sevilla, CP 41011, Spain

ARTICLE INFO

Article history:
Received 11 October 2016
Revised 18 May 2017
Accepted 19 May 2017
Available online 23 May 2017

Keywords:
Electricity consumption
Hourly demand
Load profiling
Time-series clustering
Clustering features selection
Tree classification methods

ABSTRACT

Clustering in energy markets is a top topic with high significance on expert and intelligent systems. The main impact of is paper is the proposal of a new clustering framework for the automatic classification of electricity customers' loads. An automatic selection of the clustering classification algorithm is also highlighted. Finally, new customers can be assigned to a predefined set of clusters in the classification phase. The computation time of the proposed framework is less than that of previous classification techniques, which enables the processing of a complete electric company sample in a matter of minutes on a personal computer. The high accuracy of the predicted classification results verifies the performance of the clustering technique. This classification phase is of significant assistance in interpreting the results, and the simplicity of the clustering phase is sufficient to demonstrate the quality of the complete mining framework.

1. Introduction

New technologies derived from the paradigm of Smart Grids (Tuballa & Abundo, 2016) have increased the control and monitoring of electricity consumption by customers, distribution companies, and retailers. This new scenario has led to an exponential growth in the available information concerning the grid and consumption. Thus, these technologies have led to the emergence of new services, and the increased efficiency and reliability of electricity supplies. To facilitate interaction with other systems, these new services must be able to analyse huge amounts of information in a short time (Fang et al., 2016). To achieve this goal, analysis methods and modelling designs must be constructed using big data platforms (Diamantoulakis, Kapinas, & Karagiannidis, 2015) such as Apache Hadoop (Hafen, Gibson, van Dam, & Critchlow, 2014) or Spark (Shyam, Kumar, Poornachandran, & Soman, 2015). In the current regulation model of the electricity sector, one of the main targets is to improve the performance of distribution, thus increasing the level of knowledge about demand. The most common way to evaluate energy efficiency is to evaluate the behaviour of the customers' load curve, including possible displacements in peak hours (Ferreira, de Oliveira Fontes, Cavalcante, & Marambio, 2015). Accurate knowledge of customers' consumption patterns represents a worthwhile asset for electricity providers in competitive electricity markets. Various approaches can be used to group customers that exhibit similar electricity consumption behaviour into customer classes (Chicco et al., 2004; Xu & Wunsch, 2005). Dynamic clustering can be applied (Benítez, Quijano, Díez, & Delgado, 2014; Lee, Kim, & Kim, 2011), with the focus on large-scale customers (Tsekouras, Hatziargyriou, & Dialynas, 2007; Zhang, Zhang, Lu, Feng, & Yang, 2012). The main idea is to identify customers hourly load profiles (HLPs) (Chicco, 2012; Grigoras & Scarlatache, 2014) and develop a rule set for the automatic classification of new consumers (Halkidi & Vazirgiannis, 2008; Ramos, Duarte, Duarte, & Vale, 2015). Several customer parameters, i.e. economic size, economic activity, and energy consumption, are typically used in current models (Dzobo, Alvehag, Gaunt, & Herman, 2014). In the market scenario, electricity providers have been given new degrees of freedom in defining tariff structures and rates under regulatory-imposed revenue caps (Granell, Axon, & Wallom, 2015). This requires a suitable grouping of the electricity customers into customer classes (Figueiredo, Rodrigues, Vale, & Gouveia, 2005). Other applications of load classification including the identification and correction of erroneous data and load forecasting (le Zhou, lin Yang, & Shen, 2013). Statistical techniques such as k-means (López et al., 2011), fuzzy techniques (Azadeh, Saberi, & Seraj, 2010), and frequency-domain load pattern data (Carpaneto, Chicco, Napoli, & Scutariu, 2006) have been

* Corresponding author.
E-mail addresses: fbiscarri@us.es (F. Biscarri), imonedero@us.es (I. Monedero), agarcia@us.es (A. García), juaguealo@us.es (J.I. Guerrero), cleon@us.es (C. León).

http://dx.doi.org/10.1016/j.eswa.2017.05.049
0957-4174/

used. Although different clustering methods are used for load classification (Rasanen, Voukantsis, Niska, Karatzas, & Kolehmainen, 2010), the key requirements are some load data measuring and collection platform for automated meter reading (AMR), computing software such as MATLAB, SPSS, or R, and high-performance computers. The present study was conducted using the R software.

2. Mining framework for load classification

Load classification includes a pre-clustering phase to distinguish between different categories of customers. This first categorization of customer loads considers economic reasons. During the pre-clustering phase, the main feature is the contracted tariff, which usually determines the expected load profile. Other fields of interest are the seasonal variation in electricity consumption and the individual consumer categories: households, agriculture, industry, private services, and public services. For example, agriculture consumption is not as systematic as for the other categories, and is heavily dependent on meteorological variables such as temperature, cloud cover, and daylight hours. Consumption on workdays and non-workdays differs between months and different categories, except in certain industries where the monthly profiles are assumed to be identical. Unfortunately, individual consumer categories are not always specified in the electric company database, which means that this useful information is not available for clustering purposes. After the pre-clustering phase, a data reduction process will be performed. The sampling rate of smart meters enables 24–96 consumption data per day, i.e. a sample every hour or every 15 min. This represents a significant computation time for millions of customers, each one described with a 96-dimensional vector per day. This huge quantity of data necessitates the use of data reduction and characterization techniques. Furthermore, significant information will be preserved during the reduction process. The algorithm could be run on a big data system, such as those based on Apache Hadoop or Spark libraries. The use of these infrastructures increases the efficiency and reliability of the algorithm in large-scale databases, which decreases the computation time. Different techniques for this purpose include principal component analysis (Chicco, Napoli, & Piglione, 2006), harmonic analysis (Carpaneto, Chicco, Napoli, & Scutariu, 2006), and the wavelet representation (Mallat, 1989). López et al. (2011) proposed the daily mean power values, calculated during time-of-use pricing (usually two daily periods, named peak and valley hours). This paper presents a new data characterization that reduces the computation time with respect to the techniques mentioned above. The pre-clustering phase reduces the information on each customer to a vector composed of a few features. This vector is used as the input to the clustering process. This clustering phase includes several tasks: the selection of the clustering algorithm and the optimum number of clusters. Validation techniques (Halkidi, Batistakis, & Vazirgiannis, 2001) can be applied during this step to ensure the quality of the clustering 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 method, the final partitioning of data generally requires some kind of evaluation. Thus, the main output of the clustering phase is the classification of a sample of customers into clusters. In many cases, the experts in the application area have to integrate the clustering results with other experimental evidence and analysis in order to draw appropriate conclusions. In other words, electric company experts would validate and interpret the pragmatic usefulness of the clustering. The final step is the classification phase. Classification assigns new customers to a predefined set of categories or clusters. The clustering phase produces initial categories in which the values in a dataset are classified during the classification process. The classification phase is of great assistance in interpreting the results, as we show in the following sections. The simplicity of the clustering phase convinces an electric company expert of the quality of the complete mining framework (Fig. 1).

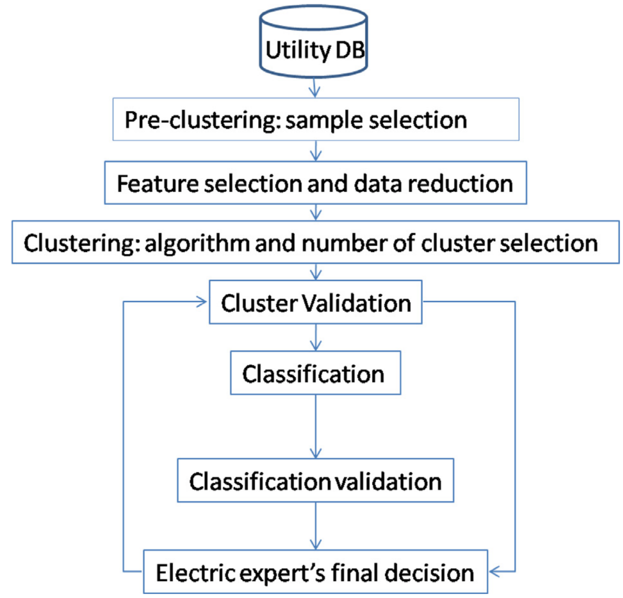

Fig. 1. Clustering and classification mining framework.

3. Pre-clustering and feature selection phases

During this phase, a sample of similar customers is selected based on their consumption and other economic criteria. The individual consumer categories are often incorrectly specified in company databases. Most customers appear as households, despite having a high contracted power. Thus, we selected a sample of customers with a certain contracted power (tariff 3.0A), a timeframe that ensures non-seasonal variations (three months), and a common climate location. These customers are located in nine adjacent villages around Seville in Andalucia, Spain. The differences between the customer environments were minimized to guarantee the homogeneity of the subsequent clustering. The sample contained a total of 218 customers. The 3.0A tariff is a time-to-use tariff for low-voltage customers (below 1 kV). There are three defined periods for pricing: peak (18–22 h), valley (0–8 h), and flat (8–18 h and 22–24 h). According to information from Spain National Commission of Energy (CNE), such customers represent approximately 2% of all electricity consumers in Spain. Feature selection and data reduction are necessary tasks. Twenty-four hourly data points per customer per day would be unmanageable in terms of computation time. Hence, researchers often use the mean daily power or the mean power during each pricing period (mean peak hours power, mean valley hours power, mean flat hours power) (López et al., 2011). Additionally, a number of studies distinguish between different months and working or non-working days (Chicco, 2012). There are two practical problems in the use of these features, one regarding the use of the mean power and another related to the use of the pricing periods. With respect to calculations based on the pricing periods, customers often vary their load profiles over the same period. Company experts prefer to divide the daytime into several periods based on the true electricity use. These periods are highly dependent on the economic activity and climate location of the consumer. Sample tests have shown that the fol-