Energy and Buildings 73 (2014) 137–145

Contents lists available at ScienceDirect

Energy and Buildings

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

Pattern recognition algorithms for electricity load curve analysis of buildings

Ioannis P. Panapakidis[a], Theofilos A. Papadopoulos[a], Georgios C. Christoforidis[b,*], Grigoris K. Papagiannis[a]

[a] Power Systems Laboratory, Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece
[b] Department of Electrical Engineering, Technological Education Institute of Western Macedonia, 50100 Kozani, Greece

ARTICLE INFO

Article history:
Received 18 November 2013
Received in revised form 2 January 2014
Accepted 5 January 2014

Keywords:
Clustering algorithms
Load profiling
Demand side management
Electricity demand analysis
Energy efficiency
Smart buildings

ABSTRACT

Buildings consume 40% of the total primary energy and 30% of the annual electricity, contributing significantly to greenhouse gas emissions. Naturally, therefore, building energy efficiency and notions like the nearly zero energy buildings are continuously gaining importance and popularity as means to reduce carbon emissions and the strong dependence on fossil fuels. A step towards this direction is the incorporation of smart grid technologies, mainly through the widespread of automatic meter reading and smart meters. This enables automatic collection of in depth information of the customer's behavior along with the building's performance and, thus, introduces new opportunities for energy saving and efficient management. However, the recorded amassing ream of data requires efficient processing and interpretation, so as to provide for meaningful information. In order to tackle this problem, this paper proposes a comprehensive methodology for the investigation of the electricity behavior of buildings, using clustering techniques. Utilizing a university campus as a case study, the proposed methodology is applied to the load curves of different buildings leading to the determination of an optimum clustering procedure. The methodology may be generalized for any type of building.

1. Introduction

The energy performance of buildings is a key focus area of the current scientific and legislation framework worldwide. Especially in Europe and over the last decade, the European Union (EU) has established several directives in order to improve the existing situation. For instance, the Recast of the Energy Performance of Buildings Directive (EPBD) [1] contains a range of provisions to improve the energy performance of buildings in order to contribute in achieving the 20-20-20 by 2020 target [2]. It introduces the notion of "nearly-zero-energy buildings" and sets this target for all new buildings by 2020 and for all new public buildings by 2018. However, the goal of reducing energy consumption is proving difficult to achieve, leading to the introduction of a special directive 2012/27/EU. This established a common framework of measures for the promotion of energy efficiency within the EU, in order to ensure the achievement of the Union's 20% target on energy efficiency to 2020 and to pave the way for further energy efficiency improvements beyond that date [3]. Public buildings are a clear focus for policy makers, since there is a significant potential of energy savings in heating, cooling and lighting especially in such buildings [4,5], resulting in an increased interest in the design, implementation and assessment of energy efficiency measures.

The legislated display energy certificates, introduced by the EPBD, constitute a first step to access the energy performance and propose effective measures, in order to implement energy saving strategies in buildings. However, these are not adequate to analyze and characterize in detail the individual components of the buildings' energy consumption as well as to identify and develop good practices and new strategies for delivering efficiency and environmental performance improvements. With the advancement of smart grid technologies, in depth information considering the customer's usage and the building's energy performance can be accessible using automatic meter reading (AMR) and smart meter systems installed in large buildings [5]. Nevertheless, the obtained amassing reams of raw data require efficient processing and interpretation to provide usable information. Therefore, sophisticated methods and load profiling tools are needed for the classification and analysis of the collected energy data and especially of electricity load curves of tertiary buildings, since in EU the tertiary sector accounts to almost 30% of the total annual electrical consumption.

* Corresponding author. Tel.: +30 24610 68141.
E-mail addresses: ipanap@ee.auth.gr (I.P. Panapakidis), thpapa@auth.gr (T.A. Papadopoulos), gchristo@teikoz.gr (G.C. Christoforidis), grigoris@eng.auth.gr (G.K. Papagiannis).

0378-7788/$ – see front matter
http://dx.doi.org/10.1016/j.enbuild.2014.01.002

Managers of buildings, campuses or other large facilities can significantly benefit from such methods and tools. Profiling and understanding the electricity consumption patterns and trends of their facilities will enhance the implementation of energy efficiency measures, electricity waste elimination and demand side management (DSM) solutions [6–8]. Another stakeholder of interest is energy service companies (ESCOs) that provide energy contracts at a broad range of comprehensive energy services, including the design and implementation of energy savings, energy conservation and energy supply solutions. The determination and implementation of such energy saving projects requires measurement and analysis of the building energy consumption profiles, using feasibility studies [9]. Distribution system operators (DSOs) may take advantage of load profiling tools as well, enabling the efficient operation and management of the distribution grid in real-time by identifying and predicting the behavior of large building customers. In general, the development of load profiling tools facilitate: (a) the comparison of similar buildings in terms of the corresponding energy consumption, (b) the establishment of benchmarking procedures, boundaries and classification schemes and (c) the proposing of possible energy efficient and environmental improvements [10,11].

In order to analyze and classify the electricity consumption behavior of a building, the use of clustering algorithms to form load curve clusters has been proposed recently [12,13]. Each cluster has a prototype load diagram, which is the load profile of the corresponding class. Common clustering algorithms in load profiling are the K-means, the minimum variance criterion (MVM), the self-organizing Map (SOM) and the Fuzzy C-means (FCM) that are mainly used due to their effectiveness [12–14]. Other less used approaches in the load curve clustering literature include the Hopfield neural network [15], the ISODATA algorithm [16] and the Support Vector Clustering (SVC) [17]. Apart from the above mentioned algorithms, hybrid combinations between them are also considered. On the other hand, classical classification (or supervised machine learning) schemes like the Support Vector Machines (SVM) [18] or the Bayes classifier [19], are mainly applied as a classification method and not as a load profiling solution. This is due to the fact that the categorization of the electricity consumption patterns of buildings is usually formulated as an unsupervised machine learning task, i.e. the number of clusters and their respective composition are a priori unknown variables. For such problems, special attention should be given to the selection and training of the appropriate pattern recognition algorithms. Additionally, several techniques have been proposed for the transformation of the load profiles to the frequency-domain, as a means to compress the size of the clustering input data set [20].

As the amount of available electricity consumption data will increase considerably in the following years due to the planned massive roll-outs of smart meters installations worldwide, the need to find scientifically valid and efficient ways to present such data in meaningful ways and in compact forms is urgent. Hence, the main aim of this manuscript is to provide recommendations to building managers and other interested stakeholders about the use of common clustering algorithms and their variants in representing the complex behavior of buildings in a compact and meaningful form. To achieve that, we use as a case study the application of clustering algorithms for the demand pattern analysis of a set of university buildings. The performance of different clustering methods, including the K-means++, SOM, MVM and the FCM combined with both time and frequency-domain data representation techniques is evaluated. The analysis is flexible and can be generalized to include all types of buildings, taking into account the exclusive energy characteristics of customers.

2. Case study – university campus

Aggregated daily electricity load data of the buildings of the Aristotle University of Thessaloniki (AUTH) Campus in Greece are considered. Twenty-seven buildings of different use are founded inside the campus and include offices, teaching rooms, lecture halls and laboratories, while one building is the sports center. The data are recorded from nine SCADA terminal units installed at the Faculty of Sciences (FoS), Engineering (FoE) with two monitoring units, Law-Economics and Political Sciences (FoLEPS), Veterinary Medicine (FoVM), Dentistry (FoD), the central library, the student club and the gym. Energy data are recorded from January 2010 to December 2011, resulting in 18 data sets of 365 patterns (the daily load curves). The daily load curve is represented by a vector of 96 elements, corresponding to the recorded average active power with a sampling rate T_{sam} equal to 15 min.

3. Proposed methodology for electricity load profile analysis

The procedure described for the buildings of the AUTH campus is general and can be applied to any type of tertiary building. In this aspect, this work provides a practical methodology for the investigation of the electricity behavior of buildings, the possible potential energy waste elimination and energy efficiency improvement by the efficient processing of load curve data. The proposed methodology is illustrated in Fig. 1 and can be summarized in the following steps.

- *Data collection:* the chronological load curves of a building are recorded with a 15 min sampling rate. Data preprocessing is applied, in case of corrupted data by noise or network faults.
- *Data transformation:* load curve data are transformed to the frequency-domain using the Fast Fourier Transform (FFT) algorithm and a set of frequency-domain features is acquired.
- *Application of the clustering algorithms:* the SOM (square)/Kmeans++ algorithm is applied to each set of features, using the proposed guidelines of this work.
- *Definition of the optimal number of clusters:* the optimal number of clusters is obtained using the knee-point criterion for the ratio of Within Cluster Sum of Squares to Between Cluster Variation (WCBCR) measure [12,21]. From a practical point of view, if this number is significantly high, the optimum number of clusters value is selected between 15 and 20, according to the performance of the WCBCR measure [14].
- *Application of DSM measures:* the characteristic load profiles of each cluster are analyzed, using a load factor term expression. The corresponding DSM measures are proposed and implemented.

4. Demand pattern modeling

Preliminary grouping of the recorded data into subsets, may increase the efficiency of the clustering procedure and reduce the size of the input data set. Therefore, the time- and frequency-domain techniques are considered for the analysis of the load data. In time-domain (TD) modeling the data set containing the metered data in kilowatts (kW) is $P = \{p_j, j = 1, \ldots, N\}$, where $N = 365$ and each daily load curve is expressed as a vector p_j:

$$p_j = [p_{j1}, \ldots, p_{j96}] \tag{1}$$

For each building, two individual data sets are considered, corresponding to the years 2010 and 2011. The load curves are grouped according to their shape similarity, since at the main clustering