Renewable and Sustainable Energy Reviews 91 (2018) 861–868

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

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

Identifying household electricity consumption patterns: A case study of Kunshan, China

Ting Yang*, Minglun Ren, Kaile Zhou

School of Management, Hefei University of Technology, Hefei 230009, China

ARTICLE INFO

Keywords:
Electricity consumption patterns
Load profiling
Smart energy management
Case study
Smart grid

ABSTRACT

A case study of residential electricity consumption patterns mining and abnormal user identification using hierarchical clustering is presented in this paper. First, based on a brief introduction of hierarchical clustering, a process model and the specific steps of electricity consumption patterns mining in smart grid environment are proposed. Then, a case study using the daily electricity consumption data of 300 residential users in an eastern city of China, Kunshan, from November 16, 2014 to December 16, 2014, is presented. Through the implementation of hierarchical clustering, 9 abnormal users and 4 types of monthly electricity consumption patterns are successfully identified. The results show that most residential users in Kunshan city, nearly 81%, have a similar monthly electricity consumption pattern. Their average daily electricity consumption is about 7.73 kWh in the early winter with small fluctuations. Also, their daily electricity consumption is significantly associated with the temperature changes. However, it is worth noting that the special electricity consumption patterns of a small proportion of electricity users cannot be ignored, which is of great significance for the planning, operation, policy formulation and decision-making of smart grid.

1. Introduction

In the Internet of Things (IoT) environment [1], with the wide application of Radio Frequency Identification (RFID), infrared sensors, Global Positioning System (GPS) and many other advanced sensing devices, the states, movement and operation related data of objects can be collected in near real time. Then based on the high-speed data processing and efficient data mining, intelligent monitoring, identification, tracking, positioning, control and management can be achieved in the IoT environment [2–4]. Smart grid [5–7] is an important application form of IoT in the electric power sector. It takes advantage of the advanced sensing, measuring and control technologies, as well as the two-way high-speed communication networks and agile decision-support systems. Thus, the safe, reliable and stable operation and the cost-effective and environmentally friendly goals of power system can be realized.

Advanced Metering Infrastructure (AMI) [8] is an advanced data collection and processing system of smart grid. It is composed of smart meters installed in the user-side, metering data management system located in power companies, and the communication system which connect both of them. Smart grid uses the AMI system to collect, measure, and store the consumers' behavioral data. Then analysis and applications can be conducted with the user generated data. For example, electricity consumers can be divided into groups with different electricity consumption behavioral characteristics using clustering algorithms [9–11]. Then the electricity consumption patterns of each group of consumers can be extracted from the data mining results, which are of great importance for the optimal operation of power system and marketing strategies development [12–14]. In addition, data preprocessing techniques can also be used for bad data identification [15], and supervised classification algorithms can be used to classify a certain electricity user into the corresponding group [16]. A lot of valuable knowledge can be discovered from the users' electricity consumption data by using advanced data analysis techniques. This knowledge plays an important role in promoting the optimal operation and enhancing the reliability of power systems, achieving personalized electricity consumption and smart energy management.

There have been considerable research efforts on the electric power load data analysis and consumers segmentation [17–21]. However, most of the existing studies have focused on the daily load profiles, and few have paid attention to the monthly electricity consumption patterns of residential users using the daily electricity consumption data within a month. Monthly electricity consumption patterns are different from the daily load patterns extracted from the hourly measured or 15-min measured load data. It is important for the medium and long term power system prediction and decision-makings. In addition, some

* Corresponding author.
E-mail address: yangtinghfut@163.com (T. Yang).

https://doi.org/10.1016/j.rser.2018.04.037
Received 13 March 2017; Received in revised form 16 October 2017; Accepted 14 April 2018
1364-0321/

studies focused on complex data mining methods, which may significantly lower the efficiency of big data processing in a smart grid environment. In addition, China is the largest developing country and one of the world's largest economies. The residential electricity users in China may have different electricity consumption patterns from the consumers in other countries.

The outlier and noisy data are prevalent exist in the complex smart grid environment. The abnormal electricity use information may be either real data generated by consumers or misleading data led by system failures and external environmental influence. So it is necessary for power companies to identify these consumers with abnormal electricity use behaviors. Specifically, the consumers that have abnormal electricity consumption patterns are regarded as abnormal users in this study. The detection and identification of abnormal users can be used for supporting many decision makings of power distribution company, including the prediction of power requirement at a specific time and some demand side management (DSM) strategies development.

The objective of this study is to explore the monthly electricity consumption patterns of residential users and attempt to identify abnormal users using the hierarchical clustering method, with daily electricity consumption data of residential users in Kunshan City, China. Finally, the monthly electricity consumption characteristics of different groups are extracted and the abnormal users are identified. The results of this study can be used for many DSM tasks in smart grid. For instance, the clustering results reveal that different groups of users have different volatility in electricity consumption profiles. The users that have high volatility in electricity consumption within a period are more tend to response to the price- and incentive-based Demand Response (DR) strategies. The results can also support the power distribution company to develop targeted DR strategies or predict electricity demand, according to the amount, trend and variance of typical electricity consumption profiles.

The remainder of this paper is organized as follows. Section 2 presents a systematic review of related works. Section 3 introduces the hierarchical clustering method, and then proposes a process model and the specific steps for the exploring of electricity consumption patterns. Results and discussions are presented in Section 4. Finally, Section 5 provides the conclusions.

2. Related works

In this section, we provide a systematic review of the related research works on household electricity consumption patterns from three dimensions, namely data, methods and applications.

2.1. Different data sources for load profiling

In smart grid environment, many different sources and different types of data have been collected and used for identifying consumption patterns and analyzing user behavior. The publicly available Irish smart meter data from the smart metering trial carried out by Commission for Energy Regulation (CER) in Ireland has been wide used in many relevant studies [22,23]. Rhodes et al. [18] presented a clustering analysis of electricity use data from 103 homes in a city of USA at seasonally-resolved timescales, and discussed the drivers for variations in their electricity use. To reduce the number of data stored for each electricity consumer, Carpaneto et al. [21] provided an frequency-domain data definition in consumer classification. High-resolution residential electricity consumption data for detached houses in Sweden were used to fit the probability distributions in the research of Munkhammar et al. [24]. Hopf et al. [25] used the electricity consumption time series recorded in 30-min intervals for household classification. Räsänen and Kolehmainen [26] focused on the time series clustering of electricity use load curves using the real hourly measured data for 1035 customers during 84 days. Pillai et al. [27] investigated the generation of load profiles using publicly available load and weather data. Based on the annual load profiles, Zhong and Tam [28] studied the extraction of characteristic attributes in frequency domain (CAFD) and the following load classification. Hopf et al. [29] used the annual electricity consumption data over one, three and five years respectively, and the customers' addresses to infer household electricity consumption characteristics. Kwac et al. [30] investigated the household electricity segmentation using the 66,434,179 electricity consumption load profiles of Pacific Gas and Electric Company (PG&E) residential consumers at 1 h intervals. Granell et al. [31] explored impact of temporal resolution on the load profiles clustering, using a data set with an 8-s sampling rate and emulates lower resolution data sets. Vercamer et al. [32] aimed to assign new electricity consumers to specific groups using the smart meter measurements in 15-min intervals of 6975 consumers in Belgium.

2.2. Clustering methods for electricity consumption patterns mining

Currently, there have been many clustering algorithms and mathematical models that were used for electricity consumption patterns mining. Fidalgo et al. [17] proposed a new clustering algorithm, which related an obtained class profile with the consumers' billing data, for load profiling. Figueiredo et al. [19] proposed an electricity user characterization framework, which included the load profiling module and the classification module. Räsänen et al. [20] proposed a self-organizing maps (SOM) based data mining technique for the analysis of customer load profile with hourly measured electricity use data of consumers in Finland. SOM was also used by Verdú et al. [33] to filter, classify, and extract patterns from distributor, commercializer, or customer electrical demand databases. For large electricity customer, Tsekouras et al. [34,35] proposed a pattern recognition methodology to determine the representative daily load profiles, and the method was used for the medium voltage customers of the Greek power system. Notaristefano et al. [36] used symbolic aggregate approximation (SAX) to reduce the features and data size of electrical load consumption data. To overcome the problem that cluster number should be predetermined in many clustering algorithm, Granell et al. [37] presented a Bayesian no-parametric model for load profiles grouping, and the so called "Chinese restaurant process" method was used to solve the model based on the Dirichlet-multinomial distribution.

K-means is a popular and efficient clustering method, which has been widely used in load profiles grouping. Considering the DSM in a smart grid environment in Brazil, Macedo et al. [38] studied the typification of load curves using k-means clustering method. Al-Wakeel et al. [39] also applied k-means method to cluster domestic smart meter measurements. A combination of k-means++ and SOM method was used for clustering of load curves of a university campus by Panapakidis et al. [40]. Kim et al. [41] developed a k-means based repeated clustering method to estimate the quarter hourly load profiles of non-AMR customers. Fuzzy c-means (FCM) is also a popular clustering method that has been widely used for electricity profiles clustering [42–44]. In addition, there were also some other methods that have been used for electricity consumption data analysis, such as Mixture Model Clustering and Markov Models [45], deep-learning based framework [46], Ant Colony clustering [47], support vector machine (SVM) based model [48], feature-selection based supervised learning method [49] and subspace projection based method [50].

2.3. Applications of electricity consumption patterns

There are also some research efforts focusing on the applications of electricity consumption patterns. Electricity consumption data clustering and pattern recognition can be used to support many decision making processes of power system [51–53]. DSM and DR are the most common application areas of electricity load profiles grouping [38,54–56]. In addition, Quilumba et al. [57] investigated the system level intraday load forecasting accuracy improvement based on the clustering of electricity consumers with similar consumption patterns.