Journal of Cleaner Production 141 (2017) 900–908

Contents lists available at ScienceDirect

Journal of Cleaner Production

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

Household monthly electricity consumption pattern mining: A fuzzy clustering-based model and a case study

Kaile Zhou [a, b, *], Shanlin Yang [a, b], Zhen Shao [a]

[a] School of Management, Hefei University of Technology, Hefei 230009, China
[b] Key Laboratory of Process Optimization and Intelligent Decision-making, Ministry of Education, Hefei 230009, China

A R T I C L E I N F O

Article history:
Received 23 May 2016
Received in revised form
22 August 2016
Accepted 20 September 2016
Available online 21 September 2016

Keywords:
Electricity consumption pattern
Households
Fuzzy clustering
Smart grid

A B S T R A C T

Household monthly electricity consumption pattern mining is to discover different energy use patterns of households in a month from their daily electricity consumption data. In this study, we develop an improved fuzzy clustering model for the monthly electricity consumption pattern mining of households. First, the background of clustering and fuzzy c-means clustering is introduced. Then a process model of household electricity consumption pattern mining and an improved fuzzy c-means clustering model are provided. Three key aspects of the improved fuzzy c-means clustering model, namely fuzzifier selection, cluster validation and searching capability optimization, are discussed. Finally, the daily electricity consumption data of 1200 households in Jiangsu Province, China, during a month from December 1, 2014 to December 31, 2014 are used in the experiment. With the proposed model, 938 valid households are successfully divided into four and six groups respectively, and the characteristics of each group are extracted. The results revealed the different electricity consumption patterns of different households and demonstrated the effectiveness of the clustering-based model. The customer segmentation based on consumption pattern mining in electric power industry is of great significance to support the development of personalized and targeted marketing strategies and the improvement of energy efficiency.

1. Introduction

To promote cleaner production and sustainable development, evaluation of energy and environment efficiency is an important research area (Chafic-Thomas Salame et al., 2016; Song and Zhou, 2015; Song et al., 2015; Song and Zheng, 2016). Households have a large potential for energy savings (Salo et al., 2016; Setlhaolo et al., 2014; Sun et al., 2014). The energy consumption of households is usually affected by a variety of factors, including internal factors (habit, attitudes, values, et al.), external factors (incentives, rewards, punishment, et al.) and interpersonal ones (norms, social comparison, et al.) (Gifford et al., 2011). Therefore, the electricity consumption patterns of different households often show great difference. Understanding their electricity consumption patterns is an effective way for residential users to change their energy use behavior and improve energy efficiency. Also, the electricity consumption patterns of households are important for utilities to develop flexible and efficient marketing strategies (Shayeghi et al., 2015).

With the increasing penetration of sensing and measurement technology, network communication technology, and intelligent control technology in energy sector, large volume of energy consumption data can be generated, collected and stored in near real time (Gurung et al., 2015; Kayastha et al., 2014). The rapid development of big data and cloud computing technology combined with advanced information and communication technologies (ICTs) make traditional energy system been digitalized (Lund et al., 2015; Shomali and Pinkse, 2016), thus form the smart energy management systems (Chai et al., 2013; Chen et al., 2011; Crossley and Beviz, 2010; Zhou et al., 2016). The digitalized smart energy system provides new opportunities for energy big data analytics. Decision support systems based on energy consumption data analysis play increasingly important roles in the operation and management of energy systems. Smart grid is a specific application form of smart energy systems, which integrates the energy flow and information flow (Cui et al., 2014). The advanced metering infrastructure (AMI) (Li et al., 2013; O'Driscoll and O'Donnell, 2013) deployed in smart grid can collect large amount of energy use data of consumers in near real time. For instance, 370 million records will be collected in a distribution network with 1 million metering devices if the data are collected once a day. If the size of each collection record is 5 Kb, the

* Corresponding author. School of Management, Hefei University of Technology, Hefei 230009, China.
E-mail addresses: zhoukaile@hfut.edu.cn, kailezhou@gmail.com (K. Zhou).

http://dx.doi.org/10.1016/j.jclepro.2016.09.165
0959-6526/

volume of these data will reach 1.82 Tb (Lavastorm, 2014). With the increasing availability of energy consumption data collected by smart meters, advanced data analysis techniques can be applied to discover valuable knowledge from the data.

Electricity consumption pattern mining is an important content of energy data analysis and knowledge discovery (Abreu et al., 2012; Andersen et al., 2013; De Silva et al., 2011). Household monthly electricity consumption pattern mining is to discover different energy use patterns of households in a month from their daily electricity consumption data using clustering methods. Fuzzy c-means (FCM) is a well-known fuzzy clustering method, which has showed superiority in electricity consumption data analysis compared with hard clustering method (Kaile et al., 2013; Prahastono et al., 2008; Tsekouras et al., 2007). However, FCM algorithm has some inherent deficiencies, which may significantly influence its performance when used for electricity consumption pattern mining. First, it is difficult to determine an appropriate value of fuzzifier which is an important parameter in FCM (Zhou et al., 2014a). Second, the number of clusters is usually unknown before clustering, while it is a necessary input parameter for FCM (Pal and Bezdek, 1995). Third, the global searching capability of traditional FCM is limited, thus it is easy to fall into local optima (Zhou and Yang, 2012).

In this study, an improved fuzzy clustering model is developed for the monthly electricity consumption pattern mining of households. The main innovations and contributions of this study mainly fall into the following three aspects. First, we propose a process model of household monthly electricity consumption pattern mining based on fuzzy clustering method. Data preparation, fuzzy clustering and results application are the three main modules in the process model. Second, an improved FCM clustering is used for the grouping of household monthly electricity consumption profiles. The fuzzy clustering model is improved in three aspects, namely fuzzifier selection, cluster validation and searching capability optimization. Third, a case study based on the proposed process model is presented. The actual electricity consumption data of 1200 households in Jiangsu Province, China during a month from December 1, 2014 to December 31, 2014 are used in the case study. We believe that the most obvious advantage of this study is that an overall solution for residential monthly electricity consumption pattern mining based on improved fuzzy clustering is proposed. The proposed process model can be used in many scenarios, including energy consumption forecasting, economic dispatch and demand side management. Another advantage of this study is that a case study of China using real-world data is presented, which demonstrates the effectiveness of the proposed model and has important practical significance.

The remainder of this paper is organized as follows. Section 2 gives the background of the fuzzy clustering method. Then, an improved FCM clustering model designed for electricity consumption pattern mining is presented in Section 3, followed by the experimental results provided in Section 4. Finally, conclusions are made in Section 5.

2. Methodology

2.1. Clustering

Clustering is an important part of data mining, pattern recognition, and statistical machine learning (Hartigan, 1975; Jain and Dubes, 1988; Jain et al., 1999). For a given data set, the objective of clustering is to partition the data into several groups, such that the data objects in the same group are as similar as possible while the data objects in different groups are dissimilar to the maximum extent. Currently, many different clustering algorithms have been proposed (Guha et al., 1998; Hartigan and Wong, 1979; Navarro et al., 1997). Due to their effectiveness in data analysis and pattern recognition, clustering methods have been widely applied in many different areas, such as intrusion detection (Liu et al., 2004), image segmentation (Wang and Pan, 2014), gene expression profiling (Alizadeh et al., 2000), to name just a few.

From the mathematical perspective, clustering process can be described as follows. For a given data set $X = \{x_1, x_2, \cdots, x_n\}$, clustering algorithms partition it into c clusters, namely $(V_1, V_2, \cdots, V_c)$. The partition matrix $U(X)$ is obtained, which can be expressed as $U(X) = [\mu_{ij}]_{c\times n}$ $(i = 1,\cdots,c, j = 1,\cdots,n)$, where μ_{ij} is the membership degree of data object x_j to cluster V_i. Generally, the clustering partition results satisfy

$$\begin{cases} \cup_{i=1}^{c} V_i = X \\ V_i \cap V_j = \varnothing, \quad i,j = 1,\cdots,c; \quad i \neq j \\ V_i \neq \varnothing, \quad i = 1,\cdots,c \end{cases} \tag{1}$$

Thus, the cluster V_i is determined by

$$\begin{cases} V_i = \{x_j \,|\, \|x_j - v_i\| \le \|x_j - v_p\|, \quad x_j \in X\}, \quad p \neq i, \quad p = 1,\cdots,c \\ v_i = \sum\limits_{x_j \in V_i} x_j \Big/ |V_i|, \quad i = 1,\cdots,c \end{cases} \tag{2}$$

where $\|\bullet\|$ denotes the similarity measure between data objects. v_i is the center of cluster V_i. $|V_i|$ represents the number of data objects in cluster V_i.

Based on the different constraints of membership degree, different kinds of clustering methods can be defined as (Sledge et al., 2010)

$$M_{\mathrm{HCM}} = \left\{ U \middle| \mu_{ij} \in \{0,1\} \quad \forall i,j; \quad 0 < \sum_{j=1}^{n} \mu_{ij} < n \quad \forall i; \quad \sum_{i=1}^{c} \mu_{ij} = 1 \quad \forall j \right\} \tag{3}$$

$$M_{\mathrm{FCM}} = \left\{ U \middle| \mu_{ij} \in [0,1] \quad \forall i,j; \quad 0 < \sum_{j=1}^{n} \mu_{ij} < n \quad \forall i; \quad \sum_{i=1}^{c} \mu_{ij} = 1 \quad \forall j \right\} \tag{4}$$

For hard clustering method (a.k.a. crisp clustering), e.g., k-means, each data object can only be divided into one group with a certain membership degree of 0 or 1. While for fuzzy clustering, e.g. fuzzy c-means (FCM), each data object can belongs to more than one groups, with a membership degree between 0 and 1 to each group.

2.2. FCM clustering

FCM is a typical fuzzy clustering method, which was first proposed by Bezdek (Bezdek et al., 1984). It starts with determining the number of clusters followed by guessing the initial cluster centers. Then, the membership degree of each data object to each group can be obtained. The cluster centers and corresponding membership degrees are updated iteratively by minimizing the objective function until it converges.