Energy Conversion and Management 171 (2018) 839–854

Contents lists available at ScienceDirect

Energy Conversion and Management

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

Association rule mining based quantitative analysis approach of household characteristics impacts on residential electricity consumption patterns

Fei Wang[a,b,*], Kangping Li[a,*], Neven Duić[c], Zengqiang Mi[a], Bri-Mathias Hodge[d], Miadreza Shafie-khah[e], João P.S. Catalão[e,f,g]

[a] State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources (North China Electric Power University), Baoding 071003, China
[b] Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
[c] University of Zagreb, Faculty of Mechanical Engineering and Navala Architecture, Ivana Lučića 5, 10000 Zagreb, Croatia
[d] National Renewable Energy Laboratory, Golden, CO 80401, USA
[e] C-MAST, University of Beira Interior, Covilhã 6201-001, Portugal
[f] INESC TEC and the Faculty of Engineering of the University of Porto, Porto 4200-465, Portugal
[g] INESC-ID, Instituto Superior Técnico, University of Lisbon, Lisbon 1049-001, Portugal

ARTICLE INFO

Keywords:
Electricity consumption pattern
Household characteristics
Association rule mining
Clustering
Apriori algorithm

ABSTRACT

The comprehensive understanding of the residential electricity consumption patterns (ECPs) and how they relate to household characteristics can contribute to energy efficiency improvement and electricity consumption reduction in the residential sector. After recognizing the limitations of current studies (i.e. unreasonable typical ECP (TECP) extraction method and the problem of multicollinearity and interpretability for regression and machine learning models), this paper proposes an association rule mining based quantitative analysis approach of household characteristics impact on residential ECPs trying to address them together. First, an adaptive density-based spatial clustering of applications with noise (DBSCAN) algorithm is utilized to create seasonal TECP of each individual customer only for weekdays. K-means is then adopted to group all the TECPs into several clusters. An enhanced Apriori algorithm is proposed to reveal the relationships between TECPs and thirty five factors covering four categories of household characteristics including dwelling characteristics, socio-demographic, appliances and heating and attitudes towards energy. Results of the case study using 3326 records containing smart metering data and survey information in Ireland suggest that socio-demographic and cooking related factors such as employment status, occupants and whether cook by electricity have strong significant associations with TECPs, while attitudes related factors almost have no effect on TECPs. The results also indicate that those households with more than one person are more likely to change ECP across seasons. The proposed approach and the findings of this study can help to support decisions about how to reduce electricity consumption and CO_2 emissions.

1. Introduction

1.1. Background and motivation

Electricity has become an increasingly important energy source for the residential sector in the past few decades. It is estimated by International Energy Agency (IEA) that the share of total electricity consumption in this sector in Organization for Economic Co-operation and Development (OECD) countries has increased from approximately 24.2% in 1974 to 31.1% in 2015 [1]. Although the energy efficiency of home appliances has been significantly improved in recent years, the average electricity consumption of household end-uses in European Union-27 Countries (EU-27) still increased by about 2.5% per year in this period [2]. Therefore, more effective and targeted measures are needed to achieve the EU 20–20–20 energy goals for energy efficiency improvement and CO_2 emissions reduction [3], which requires the comprehensive understanding of residential electricity consumption patterns (ECPs) and how they relate to household characteristics (HCs). The HCs in this paper mainly refer to the characteristics of dwelling, home appliances, occupants and their behaviors. How to identify the most significant HCs affecting the residential ECPs and reveal the complex relationship between them have become the essential problems to support decisions about how to reduce electricity consumption and CO_2 emissions.

* Corresponding authors at: 18 mail-box, North China Electric Power University, 619 North Yonghua Street, Baoding 071003, China.
E-mail address: ncepu_wangfei@sina.com (F. Wang).

https://doi.org/10.1016/j.enconman.2018.06.017
Received 1 April 2018; Received in revised form 24 May 2018; Accepted 5 June 2018
0196-8904/

Nomenclature

Abbreviations

ECP	electricity consumption pattern
TECP	typical electricity consumption pattern
DBSCAN	density-based spatial clustering of applications with noise
HC	household characteristic
ARM	association rule mining
CIE	chief income earner
POESL	proportion of energy-saving lights
PODGW	proportion of double glazed windows
NU	number
LHS	left hand side
RHS	right hand side
CC	contingency coefficient

Parameters

$p(t)$	actual active power at time t
$p^*(t)$	normalized active power at time t
ε	radius
$MinPts$	minimum number of points
$\varepsilon_{initial}$	the initial value of ε
$\Delta\varepsilon$	iterative step
χ^2	test statistic for Chi-squared test
$Sup(A \cup B)$	the support of an association rule
$Conf(A \rightarrow B)$	the confidence of an association rule
$Lift(A \rightarrow B)$	the lift of an association rule
$Imp(A \rightarrow B)$	the improvement of an association rule

Fortunately, regarded as a basic step forward to smart grid, the smart meter installations have increased worldwide in recent years [4]. For example, about 2.2 million smart meters will be installed across the country in Ireland by the end of 2020 [5]. The large-scale deployment of smart meters has enabled the accumulation and storage of electricity consumption data, which provides prerequisites for the study of understanding residential ECP and how they relate to HCs. The knowledge derived from the study can not only help to improve energy efficiency [6], but also contribute to improving tariff design [7], load forecasting and distribution network planning [8–10], and demand side management strategies [11–13].

1.2. Literature review

Clustering has been the most common technique to characterize the behaviors of electricity customers and find representative ECPs in the literature. Various algorithms have been utilized to perform ECP clustering, such as K-means, K-mediods, Fuzzy C-means, hierarchical clustering, follow the leader, ant colony clustering, self-organizing maps (SOM) and Dirichlet process mixture model [14–22]. Actually, in addition to clustering algorithm, typical electricity consumption pattern (TECP) extraction is also important for ECP clustering. As the input objects processed by the clustering algorithm, the TECP of each customer should be created before clustering. Variant TECPs extracted via different methods will inevitably lead to varied ECP clustering results. The most common method obtaining the TECP of individual customer in current studies is to calculate the average value of all the load profiles within a specific period (e.g. a month or a season) [14–16].

Studies using the combination of electricity consumption data and survey information aiming to analyze the relationships between ECPs and HCs are increasing in recent years. Rhodes et al. [23] obtained two different ECPs from 103 households via clustering first and conducted the correlation analysis between profiles and HCs using the binomial probit regression subsequently. The authors found that factors such as if someone works from home, hours of television watched per week, and education levels have significant correlations with average profile shape. This work fills the knowledge gap by identifying correlations between electric customer survey data and electricity use profiles. However, the only two clusters results drawn from a relatively small size of dataset implies that it needs to be further validated for statistically significance with some other large scale data resources. McLoughlin et al. [24] presented a clustering methodology for creating a series of representative electricity load profile classes and linked them with HCs using multi-nominal logistic regression. Beckel et al. [25] estimated the characteristics of household based on supervised machine-learning techniques using electricity consumption data. Viegas et al. [26] also proposed a machine learning based methodology for the classification of new electricity customers and discovery of the drivers of different electricity consumption profiles.

1.3. Research limitations

Understanding the specific influences of various HCs on residential ECPs is challenging because both ECP clustering and association analysis need to be considered. As the literature review shows, even though there are many studies on ECP clustering and a few other works making preliminary efforts on the association analysis between HCs and residential ECPs, some limitations of these studies can be found.

In terms of ECP clustering, the method of how to form the reasonable TECP of each individual customer in a specific period is one of the crucial steps. The TECP of each customer in a specific period should be the most representative ECP in that period, which can truly reflect customer's typical electricity consumption behavior. However, the current TECP extraction method i.e. average method usually mixes many dissimilar patterns of electricity use together and leads to an unreal reflection of how electricity is actually consumed in reality.

In terms of association relation analysis method, regression and machine learning are the two most common methods. Multicollinearity will occur in most regression models if two or more predictors are highly correlated [27]. Multicollinearity means partial coefficients vary remarkably (sometimes perhaps even change from positive to negative or conversely) while small changes occurred in predictors or datasets, which makes regression models unreliable. Regarding the machine learning methods, the results obtained by these methods can indicate how well the entire bundle of predictors predicts the response variable but are unable to provide detailed information about the cause-effect relationships between explanatory variables and explained response variable.

In terms of the completeness of study, many HCs that potentially have impacts on ECPs have not been investigated in the existing studies. For example, attitudes towards electricity consumption related factors are not included in the existing literature. Whether these factors have impacts on ECPs is still unclear. Additionally, what are the key HCs driving different ECPs is still ambiguous and the explanations of how they work mechanistically are not complete.

1.4. Contributions and paper structure

This paper aims to identify the most significant household characteristics affecting the residential electricity consumption patterns and reveal the complex relationship between them, which means it is essentially a data-mining problem instead of an optimization problem. Facing the above issues, an association rule mining (ARM) based quantitative analysis approach of HCs impact on residential ECPs is proposed to address them together in this paper.

The main contributions of this paper can be summarized as follows: