Energy Conversion and Management 165 (2018) 840–850

Contents lists available at ScienceDirect

Energy Conversion and Management

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

Clustering-based analysis for residential district heating data

Panagiota Gianniou, Xiufeng Liu*, Alfred Heller, Per Sieverts Nielsen, Carsten Rode

Technical University of Denmark, Denmark

ARTICLE INFO

Keywords:
Clustering
Load pattern
Load profiling
Load transition
Variability

ABSTRACT

The wide use of smart meters enables collection of a large amount of fine-granular time series, which can be used to improve the understanding of consumption behavior and used for consumption optimization. This paper presents a clustering-based knowledge discovery in databases method to analyze residential heating consumption data and evaluate information included in national building databases. The proposed method uses the *K*-means algorithm to segment consumption groups based on consumption intensity and representative patterns and ranks the groups according to daily consumption. This paper also examines the correlation between energy intensity and the characteristics of buildings and occupants, load profiles of households, consumption behavior changes over time, and consumption variability. The results show that the majority of the customers can be represented by fairly constant load profiles. Calendar context has an impact not only on the patterns but also on the consumption intensity and user behaviors. The variability studies show that consumption patterns are serially correlated, the customers with high energy consumption have lower variability, and the consumption is more stable over time. These findings will be valuable for district heating utilities and energy planners to optimize their operations, design demand-side management strategies, and develop targeting energy-efficiency programs or policies.

1. Introduction

Information and communications technologies (ICTs) are revolutionizing today's energy management systems. The distinct characteristic is to use smart meters to monitor energy consumption. Smart meters are the digital devices that can collect energy consumption at a fine-granular time level, typically every 15 min [1]. Smart metering systems are being widely installed globally. European countries have set the goal of converting their legacy meters to smart meters by 2020 in line with the Third Energy Package in the Electricity Directive [2] and Gas Directive [3] issued by European Commission in 2006. It is expected that 72% of European consumers will own electricity smart meters by 2020 and 40% of them will have gas meters [4]. In Denmark, more than half of current electricity customers have smart meters [5]. In recent years, several large-scale smart meter installation projects have been carried out, including the project undertaken by Enel SpA in Italy covering 30 million customers between 2000 and 2005, the Linky pilot project in France involving 300,000 customers and the national Australian initiative in Victoria covering a total of 2.6 million electricity customers. Furthermore, smart meters have been installed by Chubu Electric Power in Japan for most of their high-voltage customers, office buildings, and individual households.

The main drivers behind the employment of smart metering in different countries include load management, peak or demand reduction, fraud reduction, accurate billing and water conservation [6]. Smart meters can also be utilized to develop more accurate prediction models and detailed analyses on the drivers of building energy consumption [7]. For example, the data can reveal potentially valuable information about buildings and end users that are useful to energy management. In particular, building-related data can unveil hidden correlations between energy use and its influencing factors in buildings. Smart meter data can also help developing and applying control strategies to improve building energy performance and efficiency [8]. In addition, building energy-related information can be provided to customers [9], which can help them use demand-response techniques to reduce energy consumption, improve energy efficiency, reduce carbon emissions and improve the use of renewable energy sources. There is also an expectation that smart metering or advanced metering infrastructure can contribute to demand and cost reduction and to the adoption of domestic low-and zero-carbon technologies [6]. Therefore, they can be utilized to decrease uncertainty related to building energy performance and provide detailed information on energy monitoring. Smart meter data analyses are therefore not only of value to the research community and customers but they also help utilities to improve their meter-to-cash processes. They enable them to use advanced tariff schemes and ultimately contribute to the value of metering

* Corresponding author.
E-mail address: xiuli@dtu.dk (X. Liu).

https://doi.org/10.1016/j.enconman.2018.03.015
Received 13 December 2017; Received in revised form 5 March 2018; Accepted 7 March 2018
0196-8904/

infrastructure [10,11].

Buildings account for nearly one-third of the world's energy consumption. Among others, half of the global energy consumption of buildings comes from space heating and cooling as well as hot water [12]. In heating-dominated climates, such as Scandinavia, the main source of building energy demand is space heating. The increasing availability of smart meter data makes it possible to gain insights into heating consumption of buildings to help with energy management, such as extracting hidden temporal patterns – knowledge which can not be captured at its detailed level without the use of smart meters. Despite the promising benefits, it is still difficult to obtain reliable and detailed heating data, largely due to commercial sensitivity and privacy issues. Moreover, heating sector management is typically more challenging than other energy sectors, such as electricity, due to the high variation of production and demand. It is therefore difficult to obtain reliable data on heat production, usage patterns, and production costs. In addition, the installation of heating meters is more challenging and costly than smart electricity meters [13]. This also means that data analyses in the heating sector receive much less attention.

In this paper, a clustering-based knowledge discovery in database approach is proposed for analyzing residential heating consumption data. The objective of this study is to provide the information to district heating utilities, which they can use for optimizing their operations and for better understanding their customers. At the same time, the analysis can be used by customers to understand their consumption profiles and behaviors to improve energy efficiency. In addition, this study can help utilities and decision makers to develop energy efficiency strategies and policies, as well as provide personalized energy services to specific customer segments. Due to the introduction of renewable energy sources and the electrification of the heating sector, the Danish energy sector is in a transition period, which means that the balancing of the power grid is challenging. Energy flexibility solutions can support this transition, including demand-side management techniques that balance utility and customer needs. Identification and mapping of consumption patterns enable district heating utilities to implement new operational strategies. This study will perform a statistical clustering analysis on heating consumption data of Danish dwellings connected to a local district heating network. The data consists of the load profiles of 8293 single-family households from Aarhus, Denmark. The study clusters the customers into different groups by the *K*-means clustering algorithm, ranks the groups according to their consumption intensity, and labels them using different alphabets. An exploratory analysis is conducted on the energy consumption and socio-technical data of the dwellings, which reveals the correlation between energy consumption and the characteristics of the buildings and occupants.

This paper makes the following contributions: First, this paper presents a clustering-based approach for district heating consumption data analysis, including data preparation, clustering, and the analysis based on the clustering results. Second, this paper uses the *K*-means algorithm to segment consumption intensity groups, studies the correlation between the consumption intensity and the characteristics of buildings and occupants, and analyzes the transition of consumption behaviors over time. Third, this paper identifies representative patterns by clustering normalized daily consumption patterns and studies the variability of consumption patterns using an entropy approach. Fourth, this paper leads to a number of interesting findings from the clustering-based analysis including non-intuitive results, such as the calendar impact on consumption, the serial correlation of consumption patterns, and high consumption with a lower variability.

The remainder of this paper is organized as follows. Section 2 gives a review of related work. Section 3 presents the methods applied. Section 4 presents the data and the results of the clustering-based analysis. Section 5 discusses the significant findings of this paper and the related issues. Section 6 concludes the paper and points to the future research direction.

2. Literature review

Clustering is one of the most commonly used techniques in data mining. It is used to discover groups and identify interesting distributions in the underlying data [14]. Specifically, a clustering problem is about partitioning a given data set into groups, classes or clusters such that the data points in the same cluster are more similar compared with the data points in other clusters [14]. The goals of cluster analysis are data reduction, hypothesis development and testing and prediction identification, based on groups [15]. The main steps of clustering include feature selection, choice of the clustering algorithm, validation and interpretation of results [16]. Clustering algorithms fall into the following categories based on how clusters are defined: partitional clustering, including *K*-means and *K*-medoids methods, and hierarchical clustering, which includes density-based and grid-based clustering [17]. Clustering has been applied to different fields such as biology and web mining. Clustering techniques are being applied to the analysis of smart meter data in buildings and district heating systems, because of the rapid introduction of smart meters. Knowledge about the characteristics of clusters of consumers with similar consumption patterns can facilitate the development of novel tariff schemes and improve network management [18].

Most of the studies on clustering analysis using smart meter data are found in the electricity sector where smart meters are common [19]. In [20], the clustering methods for electricity consumption pattern recognition were assessed, including hierarchical and *K*-means algorithms. It was found that the *K*-means algorithm is more useful for segmenting customer groups, and that information hidden in the consumption patterns has a great potential to be exploited towards load control, demand response actions, real-time pricing, etc. In [21], hourly electricity consumption data from 4500 smart meters, covering all categories of Danish customers were used to identify the load profiles and classify them. It was concluded that modeling of individual load profiles should be differentiated between different categories of customers, such as between industrial and residential uses, between weekday and weekend, and between summer and winter. In [18], an automatic classification method was proposed based on self-organizing maps, which identified a set of household properties that could be deduced from electricity consumption data, such as the size of household and the income of occupants. Two methods were introduced in [22] for modeling total energy consumption of buildings with regards to predictive accuracy and cluster stability: clusterwise regression and cluster validation methods to measure stability. Clusterwise regression gave very accurate predictions but relatively unstable clusters, while *K*-means gave more stable clusters but poor predictions in several clusters. Thus, clustering methods should be carefully selected considering the main objective, whether it is accuracy or stability. In [23], a Bayesian non-parametric clustering algorithm was used to distinguish specific features between electricity consumption patterns of 219 households in UK and Bulgaria. The main advantage of the method was that it was not necessary to pre-define the number of clusters before analyzing. However, the algorithm was a bit slower in data processing than other clustering techniques. It was found that features such as nationality, family size, and type of dwelling can be successfully assigned to the members of a specific cluster.

Fewer studies have been found on data analysis in the heating sector compared with the electricity sector. In [24], the heating data from 139 single-family houses in Denmark were investigated for identifying patterns in space heating profiles using the *K*-means algorithm. It was found that the heating load profiles varied with external load conditions (high, medium and low demand periods). Two main clusters of load profiles were identified for weekdays and weekend days, one of which was quite stable and the other had a higher variation throughout the day. The latter was characterized by two peaks, one in the morning and the other in the afternoon/evening. It was concluded that building characteristics like floor area, building year and type of space heating