Applied Energy 224 (2018) 116–123

Contents lists available at ScienceDirect

Applied Energy

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

Discovering gradual patterns in building operations for improving building energy efficiency

Cheng Fan[a], Yongjun Sun[b,*], Kui Shan[c], Fu Xiao[c], Jiayuan Wang[a]

[a] Department of Construction Management and Real Estate, Shenzhen University, Shenzhen, China
[b] Division of Building Science and Technology, City University of Hong Kong, Hong Kong, China
[c] Department of Building Services Engineering, The Hong Kong Polytechnic University, Hong Kong, China

HIGHLIGHTS

- Gradual pattern mining was applied for analyzing building data for the first time.
- A generic methodology was developed to analyzing big building operational data.
- A log ratio method was proposed to improve efficiency in post-mining.
- Gradual patterns are capable of describing co-variations among variables.
- Useful knowledge has been extracted for building energy management.

ARTICLE INFO

Keywords:
Gradual pattern mining
Motif discovery
Data mining
Building operational performance
Building energy efficiency

ABSTRACT

The development of information technologies has enabled real-time monitoring and controls over building operations. Massive amounts of building operational data are being collected and available for knowledge discovery. Advanced data analytics are urgently needed to fully realize the potentials of big building operational data in enhancing building energy efficiency. The rapid development of data mining has provided powerful tools for extracting insights in various knowledge representations. Gradual pattern mining is a promising technique for discovering useful patterns from building operational data. The knowledge discovered is represented as gradual relationships, i.e., "*the more/less A*, *the more/less B*". It can bring special interests to building energy management by highlighting co-variations among numerical building variables. This study investigated the usefulness of gradual pattern mining for building energy management. A generic methodology was proposed to ensure the quality and applicability of the knowledge discovered. The methodology was validated through a case study. The results showed that the methodology could successfully extract valuable insights on building operation characteristics and provide opportunities for building energy efficiency enhancement.

1. Introduction

Adopting building automation technologies for the real-time monitoring and controls over various building services systems has become a top trend in the building sector. As reported in [1], around 22% of the energy consumed during building operations can be saved using advanced building automation technologies. Considering that building operations account for 80–90% of the energy use in building lifecycle, it is essential to improve building operational performance to achieve building sustainability [2]. The energy efficiency during building operations refers to the actual operational performance of various building services systems. It can be improved through two general approaches. The first is to adopt more energy-efficient equipment [3,4]. Common practices include replacing conventional lighting systems with LED lamps, using water-cooled chillers instead of air-cooled chillers, etc. The second is to optimize the operational performance through optimal controls, fault detection and diagnosis [5,6]. The energy saving potential can be very large due to the wide existence of improper control strategies, sensor failures, equipment degradations [7,8]. In practice, the latter approach is more feasible and economical for existing buildings, as it requires few hardware changes and initial costs. Modern buildings are being equipped with Building Automation Systems (BASs), which provides real-time monitoring and controls over different building services systems. The building operational data collected by BASs typically record the indoor and outdoor environment, power consumptions and operating conditions of different building

* Corresponding author.
E-mail address: yongjsun@cityu.edu.hk (Y. Sun).

https://doi.org/10.1016/j.apenergy.2018.04.118
Received 31 October 2017; Received in revised form 23 March 2018; Accepted 30 April 2018

services. The knowledge hidden can be very helpful for developing data-driven approaches for improving building operational performance. Previous studies mainly adopted conventional data analysis methods, such as statistics and physical principles, to utilize building operational data for predictive modeling, fault detection and diagnosis, and control optimization [9]. Despite of the encouraging research results, such data analysis methods are not suitable for discovering novel knowledge from massive data sets. For instance, previous studies mainly adopted domain expertise to select input variables for predictive modeling. Only a few variables were selected as they are considered to be closely related to the output variable based on physical reasoning [10]. Such data analysis process does not fully utilize the potential of big data, and there are chances to miss underlying relationships between outputs and other unselected variables. Existing studies, which adopted unsupervised DM as the main tools, could also suffer from this problem if only partial variables were selected/considered [11,12]. Even though manually specifying a small subset of variables to be analyzed can enhance the specificity of data analysis, it may also cost the opportunities of discovering potentially useful and unexpected knowledge. Thus, advanced methodologies should be developed to efficiently and effectively extract valuable insights from different analytical perspectives and in different knowledge representations.

Data mining (DM) is a powerful technology for discovering potentially useful knowledge from large and noisy data. It has been successfully applied in various industries, such as retails, financial services and health care [13]. DM techniques can be generally classified into supervised and unsupervised learning techniques. Supervised learning techniques can discover complex and nonlinear relationships between input and output variables. Existing studies mainly applied supervised learning techniques to perform predictive modeling for building energy consumptions, system performance indices and indoor environment [14,15]. Unsupervised learning is useful for identifying the intrinsic data structures, correlations and associations. It has been applied to identify typical building operation conditions, building occupant behaviors and interactions among building variables [16,17]. Unlike supervised learning manually selecting specified/partial data as variables, unsupervised learning can consider all data as inputs and is thus more promising in discovering novel knowledge in particular the unexpected/non-familiar one. As researchers cannot foretell what patterns or rules could be discovered by unsupervised learning from a specific data set, they may not be able to predefine the exact applications for enhancing building energy efficiency. The main reason behind is that each building has its own operating characteristics and therefore, the actual operating patterns discovered may vary case by case. Meanwhile, unsupervised learning is more suitable for analyzing massive data sets, as all variables and their underlying relationships are fully explored. Considering the rapid growth in the amount of building operational data, it could be foreseen that unsupervised data analytics will play an increasingly important role in the development of intelligent and automated building energy management systems.

Frequent pattern mining is the major area of unsupervised learning [18]. It aims to extract patterns with high occurrences from massive data sets. Various algorithms have been developed to extract frequent patterns from different types of data (e.g., transactional data, temporal data and graph data) and in different knowledge representations (e.g., frequent item sets and rules). Association rule mining (ARM) is one of the most widely used techniques for frequent pattern mining. The knowledge discovered is highly interpretable due to its rule format, i.e., $A \rightarrow B$, stating that if *Event A* happens, *Event B* will also happen. Conventional ARM algorithms, such as Apriori and FP-growth, have been applied for mining association rules in building operational data [19]. These algorithms have been used to describe static relationships among building variables [20], identify anomalies in building operations [11], and spot energy conservation opportunities [12]. There are two main limitations when applying conventional ARM algorithms to analyze building operational data. Firstly, they are only capable of extracting associations from categorical data, while most of building operational data are numerical. Data discretization can be used to transform numerical data into categorical data. However, the resulting information loss can be high due to improper discretization settings. As a result, the associations discovered may be unreliable or even meaningless. For instance, the data distribution of chilled water flow rates can be either multimodal or near-uniform depending on the water pump used (e.g., constant-speed or variable-speed pumps). Proper data discretization settings can be obtained by manually examine the data distribution and define the optimal cutoff values. Nevertheless, this process can be very time-consuming, especially when the variable number is large. To tackle this practical limitation, some studies have adopted the quantitative association rule mining (QARM) as the mining technique [21–23]. QARM adopts a data-driven approach to automatically identify the intervals for data discretization and thereby, provides great flexibility and extra insights for applications. The second limitation is that conventional ARM algorithms are not capable of extracting temporal associations, which describe the relationships among variables at different time steps. Temporal associations can be denoted as $A \xrightarrow{t} B$, indicating that if *Event A* happens, *Event B* will also happen within or after t time steps. Temporal association rule mining (TARM) can be used to discover temporal associations from time series data. The knowledge discovered has been found useful for characterizing the dynamics in building operations [24]. Currently, there is a knowledge gap between building practitioners and advanced frequent pattern mining techniques. More research efforts should be made to explore the potential of advanced frequent pattern mining techniques and their applications in building energy management.

The recent development in DM has provided powerful techniques to discover a new type of frequent patterns, i.e., gradual patterns. The knowledge discovered is expressed as gradual rules, i.e., "*the more/less A*, *the more/less B*". In essence, a gradual pattern describes the co-variations among numerical variables within a temporal space. It can simultaneously address the two above-mentioned limitations in existing studies. This study investigated the usefulness of gradual pattern mining in improving building energy efficiency. A generic methodology was proposed to ensure the quality and applicability of the knowledge discovered. The paper is organized as follows: Section 2 presents the research methodology. A case study was carried out and shown in Section 3. Applications for improving building energy efficiency are shown and discussed in Section 4. Conclusions are drawn in Section 5.

2. Research methodology

Fig. 1 depicts the research outline. The methodology was designed with four main phases to ensure the reliability and efficiency in analyzing massive building operational data. The first phase was data preprocessing, which included two tasks, i.e., (1) period estimation using spectral density estimation; (2) data transformation using symbolic aggregate approximation (SAX). Period estimation was used to identify the intrinsic periodicities in time series data, based on which subsequences are generated for in-depth analysis. The SAX method was used to transform numerical subsequences into symbols, ensuring the efficiency in motif discovery. The second phase adopted motif discovery to identify frequent subsequences in massive building operational data. Gradual pattern mining was then applied at the knowledge discovery phase to extract significant co-variations among building variables in each motif. To improve the efficiency in knowledge post-mining, a log ratio-based method was proposed to automatically identify the top-k distinguishing gradual patterns in each motif. The details of each phase are elaborated in the following subsections.

2.1. Data preprocessing

Considering the time series nature of building operational data, the