Automation in Construction 50 (2015) 81–90

Contents lists available at ScienceDirect

Automation in Construction

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

A framework for knowledge discovery in massive building automation data and its application in building diagnostics

Cheng Fan, Fu Xiao *, Chengchu Yan

Department of Building Services Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

A R T I C L E I N F O

Article history:
Received 7 July 2014
Received in revised form 18 October 2014
Accepted 11 December 2014
Available online 26 December 2014

Keywords:
Building Automation System
Data mining
Building energy performance
Building diagnostics

A B S T R A C T

Building Automation System (BAS) plays an important role in building operation nowadays. A huge amount of building operational data is stored in BAS; however, the data can seldom be effectively utilized due to the lack of powerful tools for analyzing the large data. Data mining (DM) is a promising technology for discovering knowledge hidden in large data. This paper presents a generic framework for knowledge discovery in massive BAS data using DM techniques. The framework is specifically designed considering the low quality and complexity of BAS data, the diversity of advanced DM techniques, as well as the integration of knowledge discovered by DM techniques and domain knowledge in the building field. The framework mainly consists of four phases, i.e., data exploration, data partitioning, knowledge discovery, and post-mining. The framework is applied to analyze the BAS data of the tallest building in Hong Kong. The analysis of variance (ANOVA) method is adopted to identify the most significant time variables to the aggregated power consumption. Then the clustering analysis is used to identify the typical operation patterns in terms of power consumption. Eight operation patterns have been identified and therefore the entire BAS data are partitioned into eight subsets. The quantitative association rule mining (QARM) method is adopted for knowledge discovery in each subset considering most of BAS data are numeric type. To enhance the efficiency of the post-mining phase, two indices are proposed for fast and conveniently identifying and utilizing potentially interesting rules discovered by QARM. The knowledge discovered is successfully used for understanding the building operating behaviors, identifying non-typical operating conditions and detecting faulty conditions.

1. Introduction

Modern buildings, particularly the public and commercial buildings, are equipped with Building Automation Systems (BASs) for real-time monitoring and controlling of the complicated service systems, including air conditioning, lighting, vertical transportation system, security systems, etc. BASs are the products of modern information technology, computing science and control theory. They are essentially networks of a range of hardware devices (e.g., servers, workstations, digital controllers and sensors) and software (e.g., building energy management programs and network communication protocols). A recent report showed that the potential energy savings from the adoption of advanced building automation technology might reach 22% by 2028 for the European building sector [1]. The savings are amazing considering that the building sector is responsible for approximately 32% of total final energy consumption and 40% of primary energy consumption in most countries [2]. The functionalities of BASs determine the building operational performance to a large extent. To fulfill the functions of BAS, real-time operational data are collected and stored at short intervals (from tens of seconds to several minutes) which results that a tremendous amount of building operational data is available in BASs. The amount of the BAS data keeps increasing along the building life cycle. However, the big sets of data in BASs are not fully utilized due to the lack of advanced data analysis techniques and tools. Today's BASs can only perform rather simple data analysis, such as historical data tracking, moving averages and benchmarking. In the last decade, more sophisticated tools were developed and installed in BASs owing to the fruitful research and development efforts made on advanced optimization and diagnostics of buildings [3,4]. However, those tools only take advantage of a small amount of data in BASs, and focus on the problems associated with a component or subsystem. Meanwhile, BAS data usually contains a substantial number of missing values and outliers. If those data are used in data analysis, they would ruin the analysis process and the results obtained would hardly be reliable. The building automation industry needs advanced techniques and powerful tools to analyze the massive operational data in BASs so as to understand, evaluate and improve the building operational performance.

Data mining (DM) is a promising technology, which provides new approaches to handling massive and complex data. MIT Review considered DM as one of the top 10 emerging technologies that will change the world [5]. DM has been successfully applied in various fields, such as

* Corresponding author. Tel.: +852 27664194; fax: +852 2765 7198.
E-mail address: linda.xiao@polyu.edu.hk (F. Xiao).

http://dx.doi.org/10.1016/j.autcon.2014.12.006
0926-5805/

Nomenclature

AD	Abnormality Degree
AHU	Air-Handling Units
BAS	Building Automation System
CDWP	Condenser Water Pumps
CT	Cooling Towers
ELTG	Essential Power and Lighting
MV	Mechanical Ventilation
NLTG	Normal Power and Lighting
PAU	Primary Air-handling Units
PCHWP	Primary Chilled Water Pumps
PD	Plumbing and Drainage
SCHWP	Secondary Chilled Water Pumps
SD-Lift	Standard Deviation of Lift
Temp_rtn_ch	Returned Chilled Water Temperature
Temp_sup_ch	Supplied Chilled Water Temperature
VTS	Vertical Transportation System
WCC	Water-Cooled Chillers

retails, telecommunication, and financial services [6]. DM techniques can be roughly classified into two categories, i.e., supervised learning and unsupervised learning techniques. Supervised learning aims to establish the relationship between the outputs and inputs by learning from the historical data. By contrast, unsupervised learning is not guided by an explicit mining target, and its aim is to identify the underlying and unknown data structures or associations between variables. Interests in the use of DM in the building field are increasing in recent years. DM techniques have found their strengths in three areas of the building field, i.e., prediction [7–9], fault detection and diagnosis [10–12], and control optimization [13–15]. However, the potential of DM in the knowledge discovery in massive BAS data has not been fully exploited. Previous research relied heavily on domain knowledge and mainly used supervised learning techniques. The problems were usually predefined and only a small subset of BAS data was used. For instance, in the development of the prediction model for the chiller power consumption [16], inputs to the model, e.g., the supply and return temperatures of chilled water, and the supply and return temperatures of condenser water, were selected in advance, since domain expertise tells us that these variables are the most influential variables to chiller power consumption. Even though the developed models may have higher accuracy owing to the use of domain expertise and advanced DM techniques, knowledge being discovered underlying the massive BAS data is limited.

On the one hand, although DM technology brings valuable opportunity to effective utilization of massive BAS data, the application of DM techniques in the building field faces great challenges. DM itself cannot tell the value or the significance of the knowledge discovered, and domain knowledge is still needed to interpret the knowledge for practical applications. The knowledge discovered by DM is usually enormous and may be in various forms, such as clusters, association rules, statistics, and predictive models. Meanwhile, advanced DM techniques are constantly emerging. It is not easy for building professionals to catch up the progress of DM technology. How to select the most suitable DM techniques and how to select practically valuable knowledge are two big challenges. It is not wise to attempt individual DM technique and interpret knowledge discovered on a case-by-case basis. To enable the entire building automation industry to benefit from the advanced DM technology, a generic framework for knowledge discovery in massive BAS data using DM techniques is needed. The framework should also take the poor quality of BAS data, which always contains a large number of missing values and outliers, into consideration. This paper presents a generic framework for knowledge discovery in massive BAS data using DM techniques. It is specifically designed to address all the challenges above-mentioned. The framework mainly consists of four phases, i.e., data exploration, data partitioning, knowledge discovery, and post-mining. It is expected that software tools compatible with modern BASs can be developed based on the framework. The framework is applied to analyze the BAS data of the tallest building in Hong Kong. Its values in facilitating building diagnostics are impressive.

2. Description of the framework

The framework developed is shown in Fig. 1, which mainly includes four phases. Data exploration consists of two tasks, i.e., data preprocessing and visualization. Data preprocessing aims to enhance the data quality and transform the data to suitable formats as required by DM techniques. Visualization helps the users to visually gain preliminary understanding about the data. Data partitioning aims to identify the typical building operating patterns so that the large BAS datasets can be partitioned into several subsets. It is important for enhancing the efficiency and reliability of the knowledge discovery by separately mining the data in each pattern. Knowledge discovery may adopt a number of DM techniques, such as association rule mining, clustering analysis, sequential pattern mining, ensemble learning, classification and regression, to discover the hidden knowledge. Post-mining aims to select, interpret, and make use of knowledge discovered. This study develops a novel method for selecting potentially useful association rules from a large number of rules discovered, which can significantly reduce the time needed to interpret rules using domain knowledge. Lastly, the

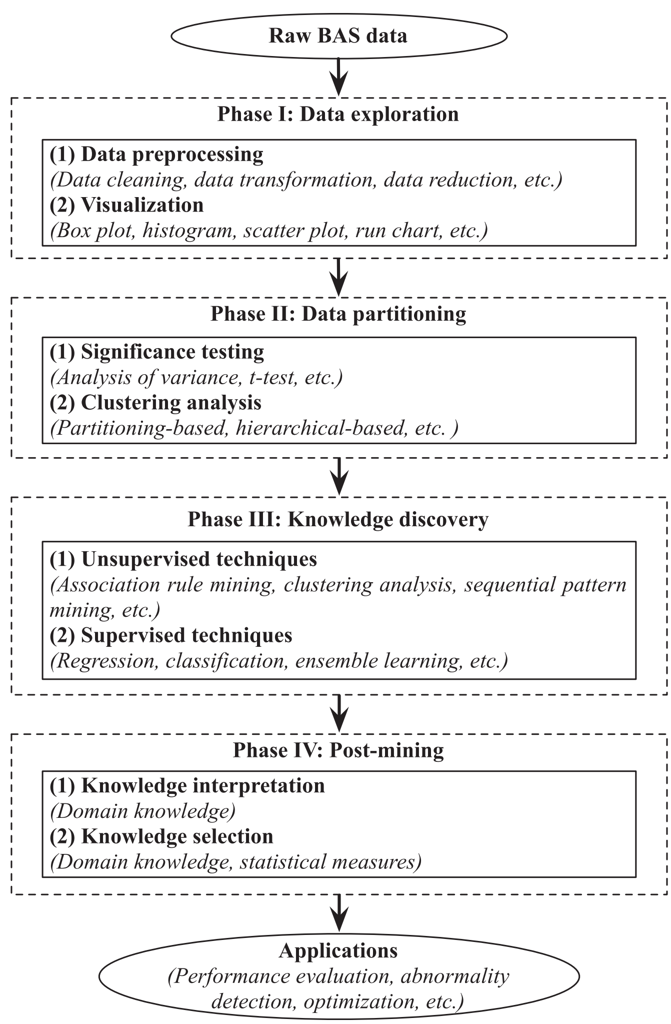

Fig. 1. Framework for mining BAS data using DM techniques.