Energy and Buildings 159 (2018) 296–308

Contents lists available at ScienceDirect

Energy and Buildings

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

Review

Unsupervised data analytics in mining big building operational data for energy efficiency enhancement: A review

Cheng Fan[a,*], Fu Xiao[b], Zhengdao Li[a,**], Jiayuan Wang[a]

[a] Department of Construction Management and Real Estate, Shenzhen University, Shenzhen, China
[b] Department of Building Services Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China

ARTICLE INFO

Article history:
Received 4 August 2017
Received in revised form 6 November 2017
Accepted 7 November 2017
Available online 9 November 2017

Keywords:
Unsupervised data mining
Big data
Building operational performance
Building energy management
Building energy efficiency

ABSTRACT

Building operations account for the largest proportion of energy use throughout the building life cycle. The energy saving potential is considerable taking into account the existence of a wide variety of building operation deficiencies. The advancement in information technologies has made modern buildings to be not only energy-intensive, but also information-intensive. Massive amounts of building operational data, which are in essence the reflection of actual building operating conditions, are available for knowledge discovery. It is very promising to extract potentially useful insights from big building operational data, based on which actionable measures for energy efficiency enhancement are devised.

Data mining is an advanced technology for analyzing big data. It consists of two main types of data analytics, i.e., supervised and unsupervised analytics. Despite of the power of supervised analytics in predictive modeling, unsupervised analytics are more practical and promising in discovering novel knowledge given limited prior knowledge. This paper provides a comprehensive review on the current utilization of unsupervised data analytics in mining massive building operational data. The commonly used unsupervised analytics are summarized according to their knowledge representations and applications. The challenges and opportunities are elaborated as guidance for future research in this multi-disciplinary field.

Contents

* Corresponding author.
** Corresponding author.
E-mail addresses: fancheng@szu.edu.cn (C. Fan), clydelee718@gmail.com (Z. Li).

https://doi.org/10.1016/j.enbuild.2017.11.008
0378-7788/

1. Introduction

The building industry has a significant impact on global sustainability. As indicated by the International Energy Agency (IEA), buildings have become the largest energy consumer in the world, accounting for over one-third of final energy consumption and an equally essential contributor of carbon dioxide emissions [1]. Building operations contribute to 80–90% of the total energy use throughout the building life cycle [2]. The energy saving potential is considerable taking into account the existence of a wide variety of building operation deficiencies. Various technologies have been developed to improve building operational performance. One prominent example is the Building Automation System (BAS), which is a network of hardware devices (e.g., servers, workstations, digital controllers and sensors) and software (e.g., building energy management programs and network communication protocols). BAS enables building operations to be more intelligent by providing real-time monitoring and controls over different building services systems. A recent report showed that the energy saving potential from the adoption of advanced BAS might reach 22% by 2028 for the European building sector [3]. A typical BAS has the ability to record a large number of measurements and control signals at short time intervals (e.g., 30-s or 1-min). As a result, massive amounts of building operational data, which are in essence the reflection of actual operating conditions, are available for data analysis. The knowledge hidden can be very helpful for a diversity of tasks in building energy management, such as predictive modeling, fault detection and diagnosis, and control optimization.

Some studies have been performed to develop advanced data analysis methodologies for mining massive building operational data. It is realized that conventional analytics, such as statistical and physical principle-based methods, are neither efficient nor effective in handling massive data sets. As a promising solution, data mining (DM) has drawn increasing attention due to its excellence in knowledge discovery from big data. DM is a multi-disciplinary subject, integrating techniques from statistics, machine learning, artificial intelligence, high performance computing and etc. There are two general types of DM, i.e., supervised and unsupervised analytics. While supervised analytics are powerful in modeling complicated relationships, unsupervised analytics are more practical and promising to discover novel knowledge given limited prior knowledge. Unsupervised analytics focus on exploring the intrinsic structures, correlations, associations and patterns in data and therefore, have the ability to discover potentially useful yet previously unknown knowledge. More importantly, the success of implementing unsupervised analytics is less dependent on domain expertise and not subject to the availability of high-quality labeled training data. It is therefore reasonable to claim that unsupervised analytics will play an essential role in the upcoming big data era of the building industry.

Compared to supervised analytics, unsupervised analytics are less known and used in the building industry. This paper presents a comprehensive review on the current utilization of advanced unsupervised analytics in mining building operational data. It aims to provide a clear picture of the status quo of unsupervised analytics for building energy management, based on which focused research can be performed in the future. The paper is organized as follows. Section 2 serves as an introduction of unsupervised analytics. Section 3 discusses the typical types of data and knowledge embedded at the building operation stage. Section 4 reviews the research and applications in mining building operational data using unsupervised analytics. Section 5 presents the challenges and possible directions for future research. Conclusion is drawn in the last section.

2. Basics of unsupervised data analytics

2.1. Supervised and unsupervised data analytics

DM analytics can be generally classified into two categories, i.e., supervised and unsupervised analytics [4]. Supervised analytics, such as boosting and bootstrap aggregating, are powerful for predictive modeling. The knowledge representations of supervised analytics are regression or classification models, which describe the quantitative or qualitative relationships between input and output variables. The success of supervised analytics is dependent on two factors, i.e., domain expertise and training data. Domain expertise is crucial for developing functional models. It is especially important for specifying model architecture, selecting model inputs, and tuning model parameters. However, the involvement of domain expertise will typically reduce the value of big data, as only a small subset of variables is used in model development. In addition, it is unlikely to discover novel knowledge, as model inputs and outputs are pre-defined. Training data refers to a set of observations where the input and output variables are both available. The quality of training data has a huge impact on the model reliability and robustness. It is worth mentioning that collecting high-quality training data can be costly, time-consuming and sometimes even not possible in practice.

By contrast, unsupervised analytics focus on discovering the intrinsic structure, correlations and associations in data. The success of implementing unsupervised analytics is not subject to the availability of training data, as there is no discrimination between inputs and outputs. The prominent advantage of unsupervised analytics is the ability to discover previously unknown knowledge [4,57]. Supervised analytics adopts a backward approach in data analysis, which means the mining target (e.g., the model output) is pre-defined. Unsupervised analytics adopts a forward approach in data analysis. All the data are taken as inputs and the mining target is not explicitly defined. The ultimate goal is to reveal interesting relationships in data, if any. In such a case, the value of big data can be best realized and the knowledge discovered might be valuable for practical applications.

This research emphasizes on the research and applications of unsupervised analytics in mining big building operational data. Considering the typical formats and types of building operational data, some of the most promising techniques are introduced as follows.