

# Extracting knowledge from building-related data — A data mining framework

Zhun (Jerry) Yu<sup>1</sup>, Benjamin C. M. Fung<sup>2</sup>, Fariborz Haghighat<sup>1</sup> (✉)

1. Department of Building, Civil and Environmental Engineering, Concordia University, Montreal, Quebec, H3G 1M8, Canada  
2. Concordia Institute for Information Systems Engineering, Concordia University, Montreal, Quebec, H3G 1M8, Canada

## Abstract

Energy management systems provide an opportunity to collect vast amounts of building-related data. The data contain abundant knowledge about the interactions between a building's energy consumption and the influencing factors. It is highly desirable that the hidden knowledge can be extracted from the data in order to help improve building energy performance. However, the data are rarely translated into useful knowledge due to their complexity and a lack of effective data analysis techniques. This paper first conducts a comprehensive review of the commonly used data analysis methods applied to building-related data. Both the strengths and weaknesses of each method are discussed. Then, the critical analysis of the previous solutions to three fundamental problems of building energy performance improvement that remain significant barriers is performed. Considering the limitations of those commonly used data analysis methods, data mining techniques are proposed as a primary tool to analyze building-related data. Moreover, a data analysis process and a data mining framework are proposed that enable building-related data to be analyzed more efficiently. The process refers to a series of sequential steps in analyzing data. The framework includes different data mining techniques and algorithms, from which a set of efficient data analysis methodologies can be developed. The applications of the process and framework to two sets of collected data demonstrate their applicability and abilities to extract useful knowledge. Particularly, four data analysis methodologies were developed to solve the three problems. For demonstration purposes, these methodologies were applied to the collected data. These methodologies are introduced in the published papers and are summarized in this paper. More extensive investigations will be performed in order to further evaluate the effectiveness of the framework.

## Keywords

building-related data,  
data mining,  
framework,  
influencing factor,  
occupant behavior,  
energy efficiency

## Article History

Received: 14 August 2012  
Revised: 10 January 2013  
Accepted: 22 January 2013

© Tsinghua University Press and  
Springer-Verlag Berlin Heidelberg  
2013

## 1 Introduction

The energy consumption in the building sector is of mounting concern. With rising living standards, building energy consumption has significantly increased over the past few decades. The current high level of consumption and the steady increase in demand for energy necessitate a thorough understanding of the major influencing factors in order to develop effective approaches to reducing building energy consumption. Factors influencing building energy consumption can be divided into seven categories (Yu et al. 2011a), as shown in Table 1.

These seven factors play an essential role in reducing

building energy consumption and efforts should be made to clearly understand their influences. However, there still are significant barriers that prevent researchers and architects from completely understanding these factors. For example, researchers and architects often observe a large discrepancy between the designed/simulated and the actual building energy consumption. The reasons for this discrepancy are not well understood and often have more to do with occupant behavior than building design. Three fundamental problems of building energy performance improvement that remain significant barriers are as follows:

(1) Building energy demand models are developed mainly for the purposes of the prediction of the total building

E-mail: haghi@bcee.concordia.ca