

# The Application of Data Mining Technology in Building Energy Consumption Data Analysis

Liang Zhao, Jili Zhang, Chongquan Zhong

**Abstract**—Energy consumption data, in particular those involving public buildings, are impacted by many factors: the building structure, climate/environmental parameters, construction, system operating condition, and user behavior patterns. Traditional methods for data analysis are insufficient. This paper delves into the data mining technology to determine its application in the analysis of building energy consumption data including energy consumption prediction, fault diagnosis, and optimal operation. Recent literature are reviewed and summarized, the problems faced by data mining technology in the area of energy consumption data analysis are enumerated, and research points for future studies are given.

**Keywords**—Data mining, data analysis, prediction, optimization, building operational performance.

## I. INTRODUCTION

ENERGY has become essential for human existence and development. The imbalance between dwindling energy supply and increasing energy demand has intensified gradually. As a result, energy conservation has become a topic that attracts global attention. Energy consumption in buildings has an enormous influence on sustainability. According to the statistics from literature, buildings account for 32% of the total amount of energy consumed worldwide. In the US and other developed nations, this figure is 41%, making the construction industry the biggest energy consumer [1]-[3]. The ratio of the buildings energy consumption approximates to one-third in some developing nations. With accelerating urbanization and higher living standards of the people in our country, the ratio of the building energy consumption will continue to rise, and may top the list of energy consumers.

The Building Energy Consumption Monitoring Platform (BECMP) and Conservation-Minded Campus Energy Conservation Supervision Platform (CCECSP) have sprung up in 2008, thanks to the substantial support from the Chinese government. According to China's report on energy saving of buildings in rural and urban areas in 2013 [4], real-time energy consumption monitoring was performed on 5,000-plus large public buildings by the end of 2013. The energy-saving campus construction projects were piloted in 210 colleges as shown in Fig. 1. The development of the system for monitoring energy consumption in buildings can provide large-scale data

information helpful in improving energy management and reducing energy consumption. It is also the most intuitive source of data concerning the operational status of buildings.

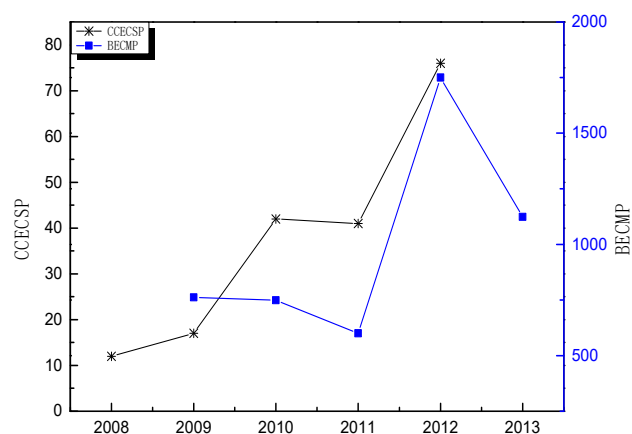


Fig. II Construction situation of BECMP and CCECSP in China

BECMP collects the temperature, humidity, pressure, electrical signals and control signals of building systems. These are critical aspects of building system operation and management, performance diagnosis, fault analysis and energy efficiency evaluation. The energy consumption data of buildings is affected by the building structure, climate environmental parameters, construction, operating modes, and user behavior patterns. Traditional methods of analysis are not sufficient in handling large-scale data which are generated during the operation of the energy consumption monitoring systems. In addition, management personnel may not be able to find and process abnormal data that may occur. This results in a serious discrepancy between the energy consumption data “on paper” and the actual energy consumption. The main reason for this is that the traditional methods adopt the energy consumption simulation scheme. This requires the users to have a good command of the building and system architecture. Furthermore, it is limited, inaccurate and cannot process useful information quickly for practical applications.

As an emerging technology, data mining has immense strength in big data processing, and it has been utilized in machine learning, human intelligence, pattern recognition, statistics, informatics, and data visualization. It has drawn growing attention from various industries and has become a bigger player in management decision-making. The data mining technology focuses on data rather than computing with complicated formulas, thereby it is easier to integrate in many applications. It is an emerging powerful technology with great

Liang Zhao is with the Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116024 China, (phone: 0086-411-84708975; fax:0086-411-84708975; e-mail: zliang@dlut.edu.cn).

Jili Zhang is with is with the Faculty of Infrastructure Engineering, Dalian University of Technology, Dalian 116024 China, (e-mail: zjldlut@dlut.edu.cn).

Chongquan Zhong is with the Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116024 China, (e-mail: zhongcq@dlut.edu.cn).

potential to discover hidden and useful knowledge in large data sets and while it has used in the construction field of the existing "rich data and poor information" dilemma [5], its application in the field of building energy systems, particularly in the energy consumption data, has been minimal.

## II. LITERATURE REVIEW

Data mining (DM) is an emerging branch of computer science that involves scouring large amounts of data in order to identify patterns and/or relationships among pieces of information in order to analyze summarize and categorize.

Seem et al. detected anomalies of energy consumption data by using the statistics-based outlier detecting method called generalized extreme studentized residual (GESR) [6]. Li et al. eliminated abnormal energy consumption data through GESR, and then classified and predicted energy consumption data via classic variable analysis (CVA). However, these methods only perform static analysis of energy consumption using historical data and thus cannot accurately detect energy consumption anomalies [7]. Qing Xiaoxia et al. mined building energy consumption data using DBSCAN clustering and decision tree-based classification methods [8]. Despite its ability to deeply understand the pattern of energy consumption in buildings, the algorithm is too complicated to be suitable for quick processing of data in the energy consumption monitoring platform. This highly technological computational process has implications for Artificial Intelligence (AI) and in recent years, DM has been gaining increasing interest in various industries, including banking, retail, healthcare, telecommunications, and even counter-terrorism [9]. The tools used in DM discover consistent patterns in the data and detect anomalies, and are also invaluable in the prediction of future trends in a market-related environment. In the construction field in particular, the use of DM procedures is eliciting promising achievements in both energy conservation and indoor environment improvement. It has been successfully used in load prediction, fault detection and diagnosis as well as optimal control. Dong et al. [10] used support vector regression models to predict monthly building energy bills. The research results validated the feasibility and applicability of this statistical method in building load forecasting. Amin-Naseri and Soroush [11] presented a hybrid neural network model combined with a clustering analysis algorithm to predict the daily electrical peak load. It was demonstrated that, compared to conventional statistical methods such as linear regression, DM-based methods achieved significant superiority in prediction accuracy. Kusiak et al. [12] developed the ensemble models of energy consumption relating to heating, ventilation, and air conditioning (HVAC) components, adopting the particle swarm optimization algorithm to identify the optimal set points of HVAC components. It was reported that 7% of HVAC energy consumption could be predicted by using their method. Ahmed et al. [13] investigated the effect of various building characteristics and climate conditions on indoor thermal comfort and on indoor illumination levels using classification techniques. Three methods, i.e. the Naïve Bayes, decision tree, and support vector machine, were applied. They indicated that

DM could be used as an effective decision aid to facilitate building operational processes. Yu et al. [14] similarly established a decision tree model for the prediction of building energy use intensity. Their research showed that the method was capable of predicting energy use intensity to a high degree of accuracy (93% for training data, and 92% for testing data). Yu et al. [15] thus employed the association rule mining in order to conserve building operational energy. The frequent-pattern growth algorithm was used to generate rules among variables of HVAC air-side systems. It was established that these rules could be used to identify energy waste, detect equipment faults, and therefore gain further insight into building operations. Cabrera and Zareipour [16] demonstrated the potential of association rule mining in identifying lighting energy waste patterns. The subsequent rules provided by the method were shown to generate significant effective energy saving measures. The simulation results achieved significant savings, as high as 70% of current energy use.

This paper analyzes the typical applications of data mining technology in recent years, presents summaries of the problems these applications are facing, and proposes research points in the future.

## III. METHODOLOGIES

This chapter summarizes the application of data mining technology in processing building energy consumption data, including the following three aspects: building energy consumption data analysis, energy consumption prediction, and data processing, which will be introduced in detail.

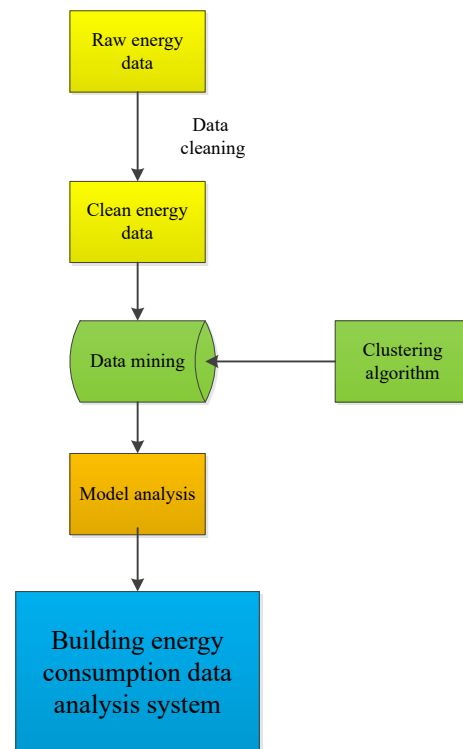


Fig. 2 Data mining process of building energy consumption platform analysis

A. Energy Data Analysis

Building energy consumption data usually has high volume and high dimension. It is very difficult to find and summarize information contained in these data. Based on these data

characteristics, this paper puts forward a clustering model for analyzing energy consumption in public buildings, as shown in Fig. 2.

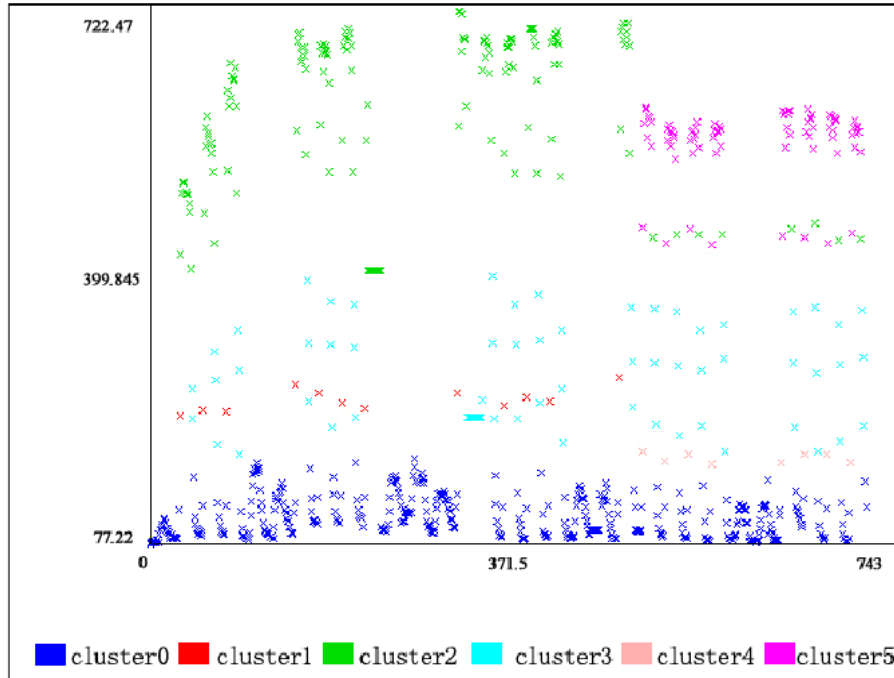


Fig. 3 The cluster distribution of an office energy consumption dataset [17]

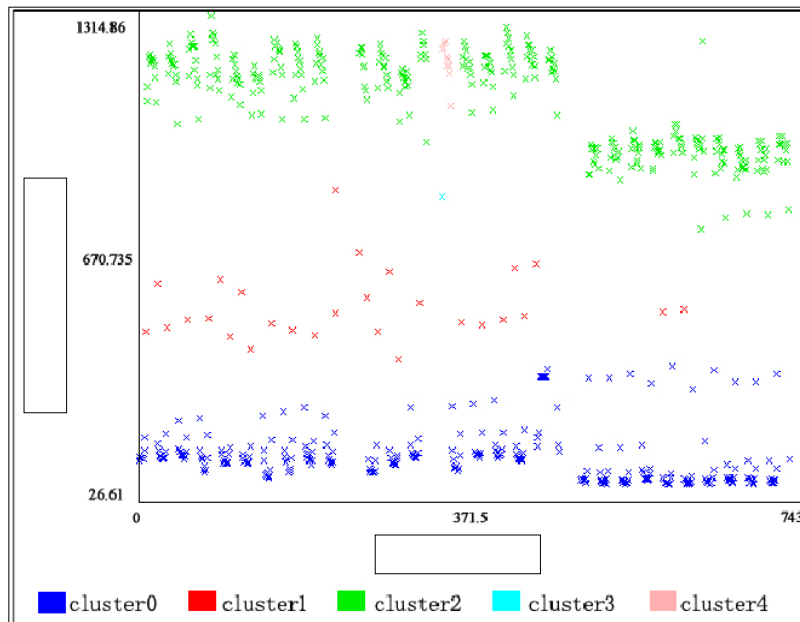


Fig. 4 The cluster distribution of a market energy consumption dataset [17]

The proposed model classifies the entire data mining process into four major stages.

1. Data collection and preprocessing: collect the data, recognize and process incorrect or incomplete data, transform the properties of energy consumption data in

public buildings, and prepare necessary background knowledge. The aim is to rid the original data of errors and transform the original data into clean data with the same format.

2. Clustering and mining: apply the clustering algorithm to the prepared data, mining knowledge that can contribute to the objectives of the energy-saving decision and energy consumption optimization.
3. Model analysis: analyze the acquired knowledge, explain their validity, and provide guidance on energy conservation.
4. Model application: experts on buildings energy saving evaluate the identified pattern, and apply this pattern to the system for analyzing energy consumption in public buildings.

Based on the hour-wise energy consumption data of an office building and a large store during a month, experiments are conducted using an improved version of Weka which incorporates the Chameleon algorithm. The proposed model is utilized to partition energy consumption in the office building and large store into three periods, shown as in Figs. 3 and 4, respectively. Energy consumption requirements are imposed on each time period to achieve energy conservation. The proposed model is also helpful in diagnosing energy consumption diagnosis in similar building types.

### B. Energy Consumption Data Prediction

Researchers have developed an effective approach to constructing the prediction models of the next-day energy consumption and peak power demand [18] which can be used to eliminate deficiencies evident in this current study. Abnormal building energy consumption profiles are first identified and are removed using feature extraction, clustering analysis, and the generalized extreme studentized deviate (GESD). Base models are then developed using eight popular predictive algorithms. A data-driven input selection algorithm called the recursive feature elimination (RFE) is applied to individually find inputs to the eight base models respectively. The flowchart of data mining methodology is shown in Fig. 5. The ensemble models are constructed by combining eight base models. A genetic algorithm (GA) is used to optimize the weights of eight base models in the final ensembles. The proposed approach is applied in the analysis of the vast energy consumption data of the tallest building in Hong Kong. The performances of individual base models and the ensemble models, together with their computation times, are then compared.

### C. Energy Data Processing

Considering the state of the data from energy consumption monitoring platform in public buildings, a hierarchical data processing method is proposed. Steps are as follows:

- (1) Data classification: Analyze and classify causes of data anomalies in the energy consumption-monitoring platform, per the building's energy consumption curves and field investigation.
- (2) Abnormal data recognition: Formulate abnormal data recognition methods based on features of various abnormal data.
- (3) Data processing: Devise abnormal data processing schemes. Delete and label different types of abnormal data.

- (4) Data mining: Utilize the algorithm to find data patterns, including the mathematical method for recognizing data anomaly, classification of buildings (devices) energy consumption modes, and the construction of sets of characteristic data concerning buildings (devices) energy consumption. This is the foundation for abnormal data recognition and processing. The proposed data processing methodology enables the energy consumption-monitoring platform to evaluate data quality and predict the energy consumption level.

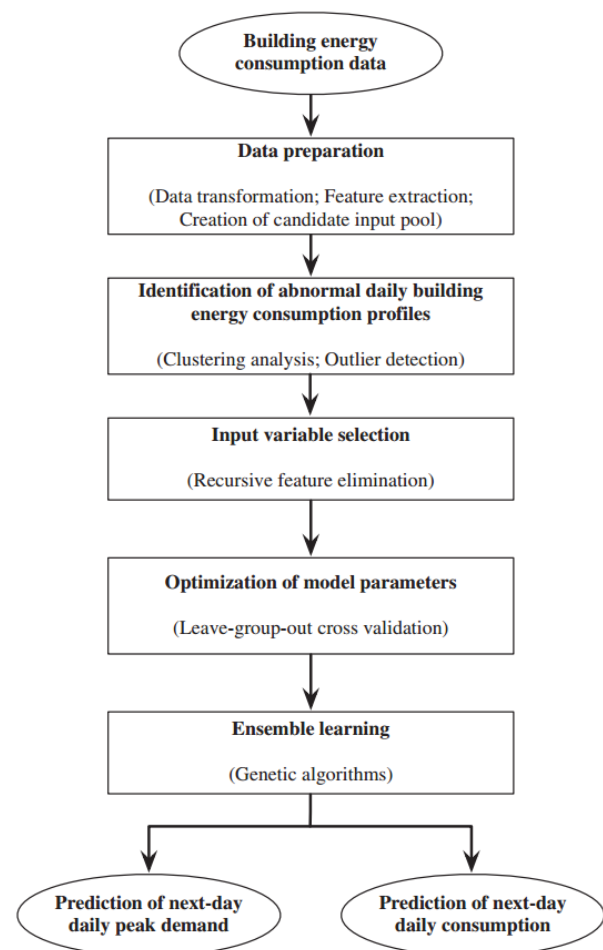


Fig. 5 Schematic outline of mining methodology [18]

Modelling energy consumption data is the first step energy conservation and optimization. The method for automatically detecting abnormal energy consumption in buildings can be further studied by using the prediction and classification results of the proposed model. Analysis of causes of abnormal energy consumption can provide insights into the energy consumed by a building, as shown in Table I. Furthermore, the proposed model can assess the influence of the building's management mode on energy consumption. It can simulate the variation of energy consumption with different factors. This is helpful in achieving energy saving by choosing better buildings

management and control mode.

Analysis of energy consumption data from the large building demonstrates the ability of the proposed method to analyze energy consumption data of each sampling point during the actual operation of energy-consuming systems, and to identify and report energy consumption anomalies promptly. It is very generic and can be used for energy-intensive places such as public buildings, sewage disposal plants and industrial manufacturers. In addition, the clustering algorithm is very noise interference-proof, making the proposed method highly robust. In practical applications, the proposed method can be embedded into the energy management system, and enable management personnel to have a quick visibility of the energy consumption situation, providing decision-making support for the formulation of energy saving measures.

TABLE I  
 DATA CLASSIFICATION OF ENERGY CONSUMPTION DATA [19]

Data Type	Data characteristic	Cause	Process method
Real data	Normal data	Reflects the actual situation	-
	Abnormal data	Reflect the actual situation	E <sub>1</sub>
Distorted data	Abnormal data	Can't reflect the actual use	E <sub>2</sub>
	Zero data	Can't reflect the actual use	E <sub>3</sub>

\* E<sub>1</sub> The user behavior is normal, but something goes wrong in equipment.  
 \* E<sub>2</sub> Sensor fault or abnormal in data transmission.  
 \* E<sub>3</sub> Sensor fault or abnormal in data transmission.

#### IV. CONCLUSION

This paper reviews the recent work on data mining technology in the application of building energy consumption data analysis, including energy consumption prediction, fault diagnosis and control optimization.

The proposed data processing method can classify abnormal data, recognize and process abnormal data hierarchically and effectively, and deepen users' understanding of data and operation of the platform for monitoring buildings energy consumption. In addition to the recognition of abnormal data, the proposed method also has the ability to intelligently compensate for the data, greatly improving data quality. The proposed data processing method forms the foundation for the functions of data quality evaluation, energy consumption prediction and warning, improving practicability of the platform for monitoring buildings energy consumption

A building's automated system is a source of data, which can be utilized to improve its energy efficiency. Despite the ability to produce reliable results, the traditional methods cannot carry out effective analysis of large-scale operational data, not to mention mining valuable information. As an emerging interdisciplinary subject, data mining is very promising for big data processing. The combination of expert data mining and large-scale buildings operation data can broaden the understanding of actual operational patterns within buildings, identify potential problems, and optimize operational performance. Currently, the application of data mining in the architecture community is still in its infancy. How to perform

data mining efficiently and apply the obtained knowledge correctly is an urgent need in this sector. This paper intends to fill that gap by utilizing data mining in the energy efficiency evaluation of buildings.

#### ACKNOWLEDGMENT

The present research is supported by the China Post-Doctoral Science Foundation (2015M571306) and the Fundamental Research Funds for the Central Universities (DUT15ZD230).

#### REFERENCES

- [1] Itard L, Meijer F, Vriens E, et al. Building renovation and modernisation in Europe: State-of-the-art review (R). Netherlands: OTB Research Institute for Housing, Urban and Mobility, 2008.
- [2] Commission E. Action plan for energy efficiency: Realising the potential (R). Brussels: Commission of the European Communities, 2006.
- [3] Zhao L, Zhang J, Liang R. Development of an energy monitoring system for large public buildings (J). *Energy and Buildings*, 2013, 66:41-48.
- [4] [http://www.gov.cn/xinwen/2014-04/25/content\\_2666605.htm](http://www.gov.cn/xinwen/2014-04/25/content_2666605.htm)
- [5] Han J W, Kamber M. *Data Mining: Concepts and Techniques* (M). 3rd ed. The Morgan Kaufmann Series in Data Management Systems, 2011.
- [6] Seem J E. Using intelligent data analysis to detect abnormal energy consumption in buildings (J). *Energy and Buildings*, 2007, 39(1): 52-58.
- [7] Li X L, Bowers C P, Schnier T. Classification of energy consumption in buildings with outlier detection (J). *IEEE Transactions on Industrial Electronics*, 2010, 57(11):3639-3644.
- [8] Qing X X, Xiao D, Wang B. A real-time monitoring method of energy consumption based on data mining (J). *Journal of Chongqing University*, 2012, 35(7):133-137.
- [9] O. Maimon, L. Rokach, *Data Mining and Knowledge Discovery Handbook*, 2<sup>nd</sup> ed., Springer, New York, 2010.
- [10] B. Dong, C. Cao, S.E. Lee, Applying support vector machines to predict building energy consumption in tropical region (J). *Energy and Buildings* 37 (2005) 545–553.
- [11] M.R. Amin-Naseri, A.R. Soroush, Combined use of unsupervised and supervised learning for daily peak load forecasting (J). *Energy Conversion and Management* 49 (2008) 1302–1308.
- [12] A. Kusiak, M.Y. Li, F. Tang, Modeling and optimization of HVAC energy consumption (J). *Applied Energy* 87 (2010) 3092–3102.
- [13] A. Ahmed, N.E. Korres, J. Ploennigs, H. Elhadi, K. Menzel, Mining building performance data for energy-efficient operation (J). *Advanced Engineering Informatics* 25 (2011) 341–354.
- [14] Z. Yu, F. Haghghat, C.M. Fung, H. Yoshino, A decision tree method for building energy demand modeling (J). *Energy and Buildings* 42 (2010) 1637–1646.
- [15] Z. Yu, F. Haghghat, C.M. Fung, L. Zhou, A novel methodology for knowledge discovery through mining associations between building operational data (J). *Energy and Buildings* 47 (2012) 430–440.
- [16] D.F.M. Cabrera, H. Zareipour, Data association mining for identifying lighting energy waste patterns in educational institutes (J). *Energy and Buildings*, 2013, 62:210-216.
- [17] Liu W F. Study of Data Mining Technique of Analyzing Energy Efficiency for Public Buildings (D). Chongqing University, China, 2010.
- [18] Fan C, Xiao F, Wang S W. Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques (J). *Apply Energy*. 2014, 127:1-10.
- [19] Yang S, Luo S X, Du M, et al. Data processing method for public building energy consumption monitoring systems based on data mining (J). *Heating ventilating & air conditioning*, 2015, 45(2):82-86.