Sustainable Cities and Society 25 (2016) 33–38

Contents lists available at ScienceDirect

Sustainable Cities and Society

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

Engineering advance

Advances and challenges in building engineering and data mining applications for energy-efficient communities☆

Zhun (Jerry) Yu[a], Fariborz Haghighat[b,*], Benjamin C.M. Fung[c]

[a] College of Civil Engineering, Hunan University, Changsha 410082, Hunan, PR China
[b] Department of Building, Civil and Environmental Engineering, Concordia University, Montreal, QC, Canada H3G 1M8
[c] School of Information Studies, McGill University, Montreal, QC, Canada H3A 1X1

ARTICLE INFO

Article history:
Available online 19 December 2015

Keywords:
Data Mining
Building Energy Use
Occupant Behavior
Big Data
Review

ABSTRACT

The rapidly growing and gigantic body of stored data in the building field, coupled with the need for data analysis, has generated an urgent need for powerful tools that can extract hidden but useful knowledge of building performance improvement from large data sets. As an emerging subfield of computer science, data mining technologies suit this need well and have been proposed for relevant knowledge discovery in the past several years. Aimed to highlight recent advances, this paper provides an overview of the studies undertaking the two main data mining tasks (i.e. predictive tasks and descriptive tasks) in the building field. Based on the overview, major challenges and future research trends are also discussed.

Contents

1. Introduction 33
2. DM applications in the building field 34
 2.1. Predictive tasks 34
  2.1.1. Building energy demand prediction 34
  2.1.2. Building occupancy and occupant behavior 34
  2.1.3. Fault detection diagnostics (FDD) for building systems 34
 2.2. Descriptive tasks 35
  2.2.1. Framework development 35
  2.2.2. Occupant behavior 35
  2.2.3. Building modeling and optimal control 35
  2.2.4. Discovering and understanding energy use patterns 36
3. Future directions and challenges—Concluding remarks 36
 Acknowledgements 37
 References 37

1. Introduction

Currently the importance of improving building energy performance for saving energy and enhancing building sustainability has been widely recognized. One effective way of achieving this objective is to uncover and extract useful knowledge from building operational data (e.g. temperature, flow rate, power and equipment states) that contains abundant valuable information on actual building performance. The widespread use of *building automation systems* (BASs) enables a tremendous amount of building operational data to be stored in building databases that also continue to expand. This rapidly growing and gigantic body of stored data, coupled with the need for data analysis, has generated an urgent need for powerful tools that can extract hidden but useful knowledge from large building databases.

As an emerging and promising technology, data mining (DM) is a powerful and versatile tool to automatically extract the valuable knowledge embedded in huge amounts of data. It can be defined in many different ways. As defined by Cabena, Hadjinian,

☆ This is an Engineering Advance paper.
* Corresponding author.
E-mail address: haghi@bcee.concordia.ca (F. Haghighat).

http://dx.doi.org/10.1016/j.scs.2015.12.001

Stadler, Verhees, and Zanasi (1998), DM is "an interdisciplinary field bringing together techniques from machine learning, pattern recognition, statistics, databases, and visualization to address the issue of information extraction from large databases." In the past several decades, researchers have been vigorously and successfully applying DM in many scientific, medical, and application domains such as banking, bioinformatics and new materials identification. Recently it has also been introduced into the building field that is a well-fit application area for DM, since it generates and collects vast amounts of data on system operation, occupant behavior, power consumption, climatic conditions and etc.

In general, DM includes six categories of widely accepted and implemented techniques (Usama, Gregory, & Smyth, 1996):

- Data classification (e.g. the decision tree method, support vector machine (SVM) and artificial neural network (ANN)),
- Clustering analysis,
- Association Rule Mining (ARM),
- Regression,
- Summarization, and
- Anomaly Detection.

Readers may refer to (Jia, Kamber, & Pei, 2012) for detailed information of these techniques. These techniques can be broadly categorized into predictive tasks and descriptive tasks (Jia et al., 2012). This paper reports an overview of the recent studies undertaking the two tasks in the building field.

2. DM applications in the building field

2.1. Predictive tasks

2.1.1. Building energy demand prediction

The prediction of building energy demand plays an important role in improving building performance. An accurate prediction needs to take various significant influencing factors of building energy demand into consideration, such as weather conditions, HVAC equipment, building envelopes and occupant behavior. The complexity and uncertainty of these factors further adds difficulties to improve prediction accuracy. DM techniques may make a breakthrough in dealing with this complexity and uncertainty.

Zhao and Magoulès (2012) indicated that DM techniques are very applicable to building energy demand prediction since they can deal with non-linear problems. ANN and SVM are the two most widely used DM techniques for this application (Ahmad et al., 2014). Kumar, Aggarwal, and Sharma (2013) applied various ANN methods, including back propagation, recurrent ANN, auto associative ANN and general regression ANN. The adopted ANN architecture significantly influences the coefficient of variation that ranges from 2% to 40%. They concluded that ANN is more suitable for the prediction of a large set of parameters than any statistical techniques. Li, Ren, and Meng (2010) reported that in many cases SVM shows higher prediction accuracy than ANN. However, training SVM can be a very slow process due to large volume of training data. The usage of parallel SVM (Zhao & Magoulès, 2011) might be a feasible alternative.

Both ANN and SVM models operate like a "black box", meaning that the model can provide a prediction but cannot provide a justification for supporting the prediction. In order to overcome this limitation, Yu, Haghighat, Fung, and Yoshino (2010) developed a building energy demand predictive model based on the decision tree method. Its competitive advantage lies in the ability to generate accurate predictive models (92% accuracy in their study) with interpretable flowchart-like tree structures that enable users to quickly extract useful information. However, the decision tree method is basically developed for predicting categorical variables other than for predicting numerical variables.

Considering that different DM techniques have their own specialties, strengths and weaknesses, it is not easy to find the best candidate for the building energy demand prediction. The recent research trend is towards the integration of different DM techniques for more accurate prediction. For example, Chou and Bui (2014) developed an ensemble model by combining ANN and SVM to predict cooling and heating demand. In order to predict the next-day energy consumption and peak power demand, Fan, Xiao, and Wang (2014) developed ensemble models by combining eight base DM models and assigning each DM modal an optimized weight based on genetic algorithm (GA). The results show that the ensemble models achieve higher prediction accuracy than those of individual base models.

2.1.2. Building occupancy and occupant behavior

Building occupancy and occupant behavior are recognized as crucial factors influencing the discrepancy between practical and simulated building energy consumption. However, it is difficult to investigate them analytically and then to develop reliable prediction models due to their complicated characteristics and stochastic nature (Yu, Haghighat, Fung, Morofsky, & Yoshino, 2011a). To meet this challenge, different stochastic models (e.g. using probabilistic methods (Sun, Yan, Hong, & Guo, 2014; Stoppel & Leite, 2014) and the Markov Chain method (Muratori, Roberts, Sioshansi, Marano, & Rizzoni, 2013)) have been proposed. However, the proposed models are severely restrained by the limitation: The modeling process tends to be complex and advanced mathematical knowledge is required.

In order to remove the above limitation, researchers have attempted to establish DM-based models. For example, Basu, Hawarah, Arghira, Joumaa, & Ploix, 2013) developed a decision-tree based model for predicting occupant behavior at the appliance level in residential buildings. D'Oca and Hong (2015) proposed a DM methodology to model office occupancy patterns and working user profiles based on big data streams. They found that the decision tree method is suitable for predicting the occupancy presence, supported by Zhao, Lasternas, Lam, Yun, and Loftness (2014), who built the occupant behavior prediction models based on appliance power consumption data in a medium-size office building. The results of both studies indicate that the modeling accuracy is very satisfactory. However, the developed decision-tree based models are static models that are hard to simulate the dynamic nature of building occupancy and occupant behavior.

The main focus of future research should be placed on testing and comparing other dynamic DM-based models and integrating them into building energy modeling programs like TRNSYS and Energy Plus. In addition, more research needs to be conducted so that architects and designers will benefit from bridging the gap between actual and predicted building energy performance.

2.1.3. Fault detection diagnostics (FDD) for building systems

Automating the process of detecting equipment and system malfunctions and making a proper diagnosis can help to ensure stable or optimal building operation. In terms of the approach to formulating the diagnostics, FDD methods can be categorized as model-based methods, which are based on prior knowledge of underlying system physics, and data-driven methods, which are based on historical data (Katipamula & Brambley, 2005).

The requirement of prior knowledge, together with complex modeling processes and heavy computational burden, imposes severe constraints on the application of the model-based methods. Comparatively, the data driven methods are much easier to use since the models are normally automatically generated. Various DM techniques have been employed as data driven methods for FDD.

For example, Magoulès, Zhao, & Elizondo, 2013) developed an ANN model based on the recursive deterministic perceptron (RDP) ANN to implement FDD at the whole building level. Moreover, they proposed a new fault diagnostic procedure for identifying and ranking abnormal equipment in the order of fault risk. Other DM techniques have also been employed for FDD such as cluster analysis (Khan, Capozzoli, Corgnati, & Cerquitelli, 2013), fuzzy logic (Lauro et al., 2014) and SVM (Mulumba, Afshari, Yan, Shen, & Norford, 2015).

In order to take advantage of different DM techniques and achieve more robust models, hybrid approaches have been proposed for FDD such as the combination of the SVM and autoregressive model with exogenous inputs (ARX) (Yan, Shen, Mulumba, & Afshari, 2014) and the combination of ANN and subtractive clustering analysis (Du, Fan, Jin, & Chi, 2014). Experimental results suggest these hybrid approaches have higher prediction accuracy and lower false alarm rates than the application of single models. However, the above DM-based models, both individual and hybrid, focused on historical data and cannot be used for real-time FDD.

Recently, researchers have shown interest in developing online FDD tools that can provide real-time or near-real-time updates on model parameters and predictions (Bonvini, Sohn, Granderson, Wetter, & Piette, 2014). DM techniques, particularly the SVM, have been utilized in these studies and shown great potential for real-time FDD since they prevent performing numerous time-consuming simulations (Yan et al., 2014). Future research should focus on verifying prediction of models based on a real world data set.

2.2. Descriptive tasks

2.2.1. Framework development

The diversity of DM techniques and their functionalities necessitates establishing and providing a framework for users who may not have much knowledge and experience with the application of DM. Basically, such framework needs to specify which DM technique/algorithm can be used for building application and which kind of pattern can be mined correspondingly. The ultimate objective is to help users to develop a set of data analysis methodologies so that they can solve the problems based on the framework. Moreover, case studies also need to be conducted in order to demonstrate its effectiveness. The framework, together with the case studies, enables users to analyze building-related data more efficiently.

Both general and specific DM frameworks have been proposed in the past several years (Yu, Fung, & Haghighat, 2013). General frameworks provide a DM technique pool and different patterns will be found given different DM techniques are selected and combined. Specific frameworks are targeted at exploring a specific pattern such as occupancy patterns and designated DM techniques are used.

Regarding general frameworks, Yu et al. (2013) established a four-component framework consisting of DM techniques/algorithms. They also proposed a step-by-step data analysis process that starts from problem definition to knowledge discovery. Fan, Xiao, and Yan (2015) developed a framework that mainly includes four phases: data exploration, data partitioning, knowledge discovery and post mining. Regarding specific frameworks, D'Oca and Hong (2015) developed a framework to discover occupancy patterns and working user profiles in office spaces. The decision-tree method and cluster analysis were combined to provide insights into patterns of occupancy schedules. However, these frameworks involve limited DM techniques and so far relatively very few problems has been tackled with the involved DM techniques.

DM frameworks can deliver a set of data analysis methodologies to practical operation, and thus efficiently generalizing the usage of DM techniques to a wider population. They provide an overview and a standardized process of the usage of DM techniques. Therefore, more frameworks, particularly specific frameworks, and case studies with high replication potential should be developed and applied for different applications.

2.2.2. Occupant behavior

DM techniques have also been employed to conduct descriptive tasks for occupant behavior in several studies. These studies focus on either one type of specific behavior in office buildings or full patterns of behavior in residential buildings. Their main objectives are to discover behavior patterns and to identify and improve the effect of occupant behavior on building energy consumption. The obtained knowledge can be integrated into building energy simulation programs especially when setting up behavior schedules.

D'Oca and Hong (2014) combined clustering analysis with ARM in order to discover occupants' window opening/closing behavior patterns in a natural ventilated office building. Based on cluster analysis, Yu, Fung, Haghighat, Yoshino, and Morofsky (2011b) developed a new methodology for examining the influences of occupant behavior on building energy consumption while removing the effect of other factors such as climate and building envelopes. Based on cluster analysis, classification analysis, and ARM, Yu et al. (2011a) developed another methodology for identifying energy-inefficient behavior in residential buildings and providing occupants with feasible recommendations to improve them. It is worth mentioning that the effective usage of DM techniques, particularly cluster analysis, for describing occupant behavior is heavily dependent on the size of established databases, which normally are restricted by various factors such as budgets and effort in practice.

Due to its complexity and randomness, the interactions between occupant behavior and other influencing factors are still poorly understood. Further studies on the DM-based methodologies are actively encouraged, and the combination of cluster analysis with ARM deserves extra attention.

2.2.3. Building modeling and optimal control

Building operational data contains valuable information that can be extracted for modeling and identifying the hidden relationship between different variables, thereby improving building operation. Also, coupling the information into building control systems can assist in optimizing building performance. A number of studies have already been conducted on such information extraction.

With the goal of improving building performance, DM techniques have been employed to model different performance variables such as furnaces' cycle idle time (Li, Miao, & Shi, 2014), thermodynamic properties of refrigerants (Küçüksille, Selbaş, & Şencan, 2011), appliance usage (Basu, Debusschere, & Bacha, 2012) and coefficient of performance for refrigeration equipment (Chou, Hsu, & Lin, 2014). For example, Ahmed, Korres, Ploennigs, Elhadi, and Menzel (2011) used Naïve Bayes, decision tree and SVM to simplify the procedure to establish models for predicting internal natural lighting and thermal comfort. The results show the high prediction accuracy and reliability of these techniques. Based on the *BAS* database of the tallest building in Hong Kong, Xiao and Fan (2014) first applied clustering analysis to identify the typical building energy consumption patterns, and then applied ARM to each cluster for discovering the hidden associations among power consumptions of major equipment such as chillers and pumps. As a result, deficit flow in the chilled water system was found and abnormal operation of the primary and secondary pumps was detected.

DM techniques have also been employed for HVAC system optimal control with the aim of minimizing the energy consumption while maintaining indoor comfort levels. Generally these techniques are integrated into controllers in order to take advantage of their strengths. For example, based on the fuzzy logic method, ANN and Genetic Algorithm, fuzzy logic controllers (Hussain, Gabbar, Bondarenko, Musharavati, & Pokharel, 2014), ANN controllers (Moon, Lee, & Kim, 2014) and GA-based controllers (Congradac & Kulic, 2009) were developed. These controllers have already been widely applied in the building field and a large number of studies can be found in literature that is not listed in this paper.

Existing studies have already demonstrated the capabilities of DM techniques as a means to effectively help improving and optimizing building performance. Potential future research includes developing more accurate and robust models, particularly a combination of mining techniques, that can be readily usable. Also, more practical applications of DM-based controllers need to be conducted in the building field in order to evaluate their usability.

2.2.4. Discovering and understanding energy use patterns

Building energy consumption patterns, both frequent and infrequent, usually can be used to help identify operation abnormalities and faults. Effective pattern recognition and corresponding knowledge extraction by mining building historical data or simulation data can lead to energy savings and performance improvement. Relevant studies at both the whole building level and sub-system level have been conducted in recent years.

At the whole building level, Miller, Nagy, and Schlueter (2015) applied clustering analysis to aggregate frequent energy use patterns so as to generate profiles that may benefit future simulation model calibration. Lange, Rodriguez, Puech, and Vasques (2013) developed a 3D visualization tool that can be integrated into BASs to visualize data in an understandable way. The work is part of the RIDER project that composed of a set of DM techniques aimed at gaining insight into building energy use patterns. Based on ARM, Yu, Haghighat, Fung, and Zhou (2012) developed a methodology for examining all associations and correlations between building operational data, thereby pinpointing energy waste as well as equipment faults. By mining simulation data, Kim, Stumpf, and Kim (2011) developed an overall data analysis process to find important patterns in order to improve the energy efficiency of building design during the early design phase.

At the sub-system level, an increasing number of studies (Lai, Lai, Huang, & Chao, 2013; Beckel, Sadamori, Staake, & Santini, 2014; Rollins & Banerjee, 2014) focused on the investigations on appliance usage patterns. Chicco (2012) provided an overview of the previous literature on clustering analysis for grouping and mining electrical load patterns, particularly residential demand profiles (Rhodes, Cole, Upshaw, Edgar, & Webber, 2014). Few studies further focused on more specific appliance usage patterns such as lighting (Motta Cabrera & Zareipour, 2013).

Patterns extracted from building historic data and simulation data provide a deep insight into the way building occupants consume energy. Utilities, policy-makers, building owners, designers and operators may benefit greatly from these extracted knowledge. Most existing studies are still focused on general pattern analysis over a long time length, limited by low resolution data available. In order to discover in-depth knowledge of building performance, the focus of future studies needs to be placed on more detailed patterns based on high resolution data such as 1-min data.

3. Future directions and challenges—Concluding remarks

This paper reviews the advances in the field of building engineering and DM applications over the last several years. It has been shown that DM techniques are an effective tool for extracting hidden but useful knowledge from building-related data, thereby improving building performance. However, some important problems limiting the applications still need to be addressed and more effort is needed. The following described the encountered and anticipated challenges, potential future developments and recommendations for the application of DM techniques to the building field:

(1) Low-quality data leads to low-quality data mining results. A key challenge is how to improve data quality in terms of reliability, completeness, consistency and resolution. There are two major reasons for the lack of high-quality data: the lack of widely-accepted benchmarks for identifying the parameters to be monitored and the lack of adequate sensors for collecting reliable and high resolution data that tend to be costly. Moreover, for residential buildings, the energy consumption data are collected from the monthly or annual bills from utility companies. These highly summarized data are insufficient to conduct detailed data analysis. The fact that end-use data on household appliance usage is seldom monitored adds great difficulties to in-depth data analysis.

(2) Wide and diverse applications have been found and various DM techniques have been employed. However, considering that DM is a relatively young and fast-growing discipline and new techniques emerges, clearly the potential of DM to extract useful knowledge from building-related data has not been fully exploited. In this view, building researchers and engineers need to keep an eye on the current trend towards the development of DM techniques, such as scalable and constraint-based mining methods, as well as their applications.

(3) Previous studies indicate that the applications have several important limitations. For example, although a deeper understanding of the data can be gained, the process of model construction and knowledge extraction relies heavily on users' prior domain knowledge (e.g. setting a threshold for ARM and determining the optimal number of clusters). Also, it is not intuitive on how to utilize the discovered knowledge. More effort is needed for the exploration of improving the efficiency and ease of the mining process and result analysis. One possible direction to take is the introduction of "action mining" or "profit mining" that has shown some success in the business world. Their DM results are not just a bunch of rules, but some suggested actions to maximize the profit or revenue. Similar concepts may be applicable in the building field.

(4) The increasing volume, velocity and variety of building-related data usually result in databases with sizes and complexity beyond the ability of traditional statistical analysis methods to process them efficiently. Such databases can be considered as "Big Data" that is a current topic under discussion in Information, Communication and Technology (ICT) in buildings' experts. Based on the definition and existing application of DM to building-related data, DM is expected to play an important role in establishing effective "Big Data" management approaches. From this perspective, it is highly recommended that future research pay more attention to the following points:
   - Passing the information on the usefulness of DM applications in the building field to policy makers and building owners so that they will, in turn, be supportive for more applications;
   - formulating acceptable benchmarks for identifying the parameters to be monitored;
   - lowering the cost of monitoring and accessing building-related data;
   - developing methodologies based on new DM techniques for diverse applications;
   - improving the presentation and visualization of DM results; and
   - investigating multiple building blocks and communities besides individual buildings.

Acknowledgements

This work was financially supported by the National Natural Science Foundation of China (No. 51408205), the Fundamental Research Funds for the Central Universities, Hunan Provincial Science and Technology Department Major Project of China (No. 2011FJ1007) and was also funded by ELS.

References

Ahmad, A. S., Hassan, M. Y., Abdullah, M. P., Rahman, H. A., Hussin, F., Abdullah, H., et al. (2014). A review on applications of ANN and SVM for building electrical energy consumption forecasting. *Renewable and Sustainable Energy Reviews*, *33*, 102–109.
This paper provides an overview of the building electrical energy prediction method using SVM and ANN. It suggests that combining multiple classification models ensure better forecasting performance. The proposed approach is promising.
Ahmed, A., Korres, N. E., Ploennigs, J., Elhadi, H., & Menzel, K. (2011). Mining building performance data for energy-efficient operation. *Advanced Engineering Informatics*, *25*, 341–354.
Basu, K., Debusschere, V., & Bacha, S. (2012). Appliance usage prediction using a time series based classification approach. In *IECON 2012—38th Annual Conference on IEEE Industrial Electronics Society Montreal, Canada*.
Basu, K., Hawarah, L., Arghira, N., Joumaa, H., & Ploix, S. (2013). A prediction system for home appliance usage. *Energy and Buildings*, *67*, 668–679.
Beckel, C., Sadamori, L., Staake, T., & Santini, S. (2014). Revealing household characteristics from smart meter data. *Energy*, *78*, 397–410.
Bonvini, M., Sohn, M. D., Granderson, J., Wetter, M., & Piette, M. A. (2014). Robust on-line fault detection diagnosis for HVAC components based on nonlinear state estimation techniques. *Applied Energy*, *124*, 156–166.
Cabena, P., Hadjinian, P., Stadler, R., Verhees, J., & Zanasi, A. (1998). *Discovering data mining: From concept to implementation*. Upper Saddle River, NJ: Prentice Hall.
Chicco, G. (2012). Overview and performance assessment of the clustering methods for electrical load pattern grouping. *Energy*, *42*, 68–80.
Chou, J.-S., & Bui, D.-K. (2014). Modeling heating and cooling loads by artificial intelligence for energy-efficient building design. *Energy and Buildings*, *82*, 437–446.
Chou, J.-S., Hsu, Y.-C., & Lin, L.-T. (2014). Smart meter monitoring and data mining techniques for predicting refrigeration system performance. *Expert Systems with Applications*, *41*, 2144–2156.
Congradac, V., & Kulic, F. (2009). HVAC system optimization with CO_2 concentration control using genetic algorithms. *Energy and Buildings*, *41*, 571–577.
D'Oca, S., & Hong, T. (2014). A data-mining approach to discover patterns of window opening and closing behavior in offices. *Building and Environment*, *82*, 726–739.
The authors first implemented clustering analysis to the measured data in different offices for obtaining patterns including motivations, duration and number of opening and typical degree of opening positions. Based on the obtained information, ARM was then conducted and consequently the occupants were divided into two typical working user profiles: physical environmental driven and contextual driven.
D'Oca, S., & Hong, T. (2015). Occupancy schedules learning process through a data mining framework. *Energy and Buildings*, *88*, 395–408.
Du, Z., Fan, B., Jin, X., & Chi, J. (2014). Fault detection and diagnosis for buildings and HVAC systems using combined neural networks and subtractive clustering analysis. *Building and Environment*, *73*, 1–11.
Fan, C., Xiao, F., & Wang, S. (2014). Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. *Applied Energy*, *127*, 1–10.
Fan, C., Xiao, F., & Yan, C. (2015). A framework for knowledge discovery in massive building automation data and its application in building diagnostics. *Automation in Construction*, *50*, 81–90.
Hussain, S., Gabbar, H. A., Bondarenko, D., Musharavati, F., & Pokharel, S. (2014). Comfort-based fuzzy control optimization for energy conservation in HVAC systems. *Control Engineering Practice*, *32*, 172–182.
Jia, Wei Han, Kamber, M., & Pei, J. (2012). *Data mining concepts and techniques* (3rd ed.). Waltham, US: Morgan Kaufmann.
Katipamula, S., & Brambley, M. R. (2005). Methods for fault detection, diagnostics, and prognostics for building systems—A review, part I. *HVAC&R Research*, *11*, 3–25.
Khan, I., Capozzoli, A., Corgnati, S. P., & Cerquitelli, T. (2013). Fault detection analysis of building energy consumption using data mining techniques. *Energy Procedia*, *42*, 557–566.
Kim, H., Stumpf, A., & Kim, W. (2011). Analysis of an energy efficient building design through data mining approach. *Automation in Construction*, *20*, 37–43.
Küçüksille, E. U., Selbaş, R., & Şencan, A. (2011). Prediction of thermodynamic properties of refrigerants using data mining. *Energy Conversion and Management*, *52*, 836–848.
Kumar, R., Aggarwal, R. K., & Sharma, J. D. (2013). Energy analysis of a building using artificial neural network: A review. *Energy and Buildings*, *65*, 352–358.
Lai, Y.-X., Lai, C.-F., Huang, Y.-M., & Chao, H.-C. (2013). Multi-appliance recognition system with hybrid SVM/GMM classifier in ubiquitous smart home. *Information Sciences*, *230*, 39–55.
Lange, B., Rodriguez, N., Puech, W., & Vasques, X. (2013). Discovering unexpected information using a building energy visualization tool. In *Proc. SPIE 8650, Three-Dimensional Image Processing (3DIP) and Applications*.
Lauro, F., Moretti, F., Capozzoli, A., Khan, I., Pizzuti, S., Macas, M., et al. (2014). Building Fan Coil Electric Consumption Analysis with Fuzzy Approaches for Fault Detection and Diagnosis. *Energy Procedia*, *62*, 411–420.
Li, Q., Ren, P., & Meng, Q. (2010). Prediction model of annual energy consumption of residential buildings. In *Proceedings of 2010 international conference on advances in energy engineering*.
Li, M., Miao, L., & Shi, J. (2014). Analyzing heating equipment's operations based on measured data. *Energy and Buildings*, *82*, 47–56.
Magoulès, F., Zhao, H.-x., & Elizondo, D. (2013). Development of an RDP neural network for building energy consumption fault detection and diagnosis. *Energy and Buildings*, *62*, 133–138.
The authors simulated the daily electricity consumption of different electric equipment in an office building over a year. The parameters of the equipment are set to default and changed values, indicating normal and faulty conditions, respectively. The generated normal and faulty datasets were then used to train and test the fault detection ability of the developed ANN model. Experimental results demonstrate that this model is highly accurate in fault detection.
Miller, C., Nagy, Z., & Schlueter, A. (2015). Automated daily pattern filtering of measured building performance data. *Automation in Construction*, *49*(Part A), 1–17.
Moon, J. W., Lee, J.-H., & Kim, S. (2014). Application of control logic for optimum indoor thermal environment in buildings with double skin envelope systems. *Energy and Buildings*, *85*, 59–71.
Motta Cabrera, D. F., & Zareipour, H. (2013). Data association mining for identifying lighting energy waste patterns in educational institutes. *Energy and Buildings*, *62*, 210–216.
This study proposed an ARM-based method to quantify and understand lighting energy waste patterns associated with occupancy patterns in classrooms in post-secondary educational institutes. Simulation results show that up to about 40% savings could be achieved by implementing the discovered rules.
Mulumba, T., Afshari, A., Yan, K., Shen, W., & Norford, L. K. (2015). Robust model-based fault diagnosis for air handling units. *Energy and Buildings*, *86*, 698–707.
Muratori, M., Roberts, M. C., Sioshansi, R., Marano, V., & Rizzoni, G. (2013). A highly resolved modeling technique to simulate residential power demand. *Applied Energy*, *107*, 465–473.
Rhodes, J. D., Cole, W. J., Upshaw, C. R., Edgar, T. F., & Webber, M. E. (2014). Clustering analysis of residential electricity demand profiles. *Applied Energy*, *135*, 461–471.
This study clustered 103 homes in Austin by their hourly electricity use patterns for each season. The optimal number of clusters was determined to represent residential demand profiles. The results will benefit utilities for formulating policies on load shifting and efficiency upgrade rebates.
Rollins, S., & Banerjee, N. (2014). Using rule mining to understand appliance energy consumption patterns. In *2014 IEEE International Conference on Pervasive Computing and Communications (PerCom)*.
Stoppel, C. M., & Leite, F. (2014). Integrating probabilistic methods for describing occupant presence with building energy simulation models. *Energy and Buildings*, *68*(Part A), 99–107.
Sun, K., Yan, D., Hong, T., & Guo, S. (2014). Stochastic modeling of overtime occupancy and its application in building energy simulation and calibration. *Building and Environment*, *79*, 1–12.
Usama, Fayyad, Gregory, Piatetsky-Shapiro, & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, *17*, 37–54.
Xiao, F., & Fan, C. (2014). Data mining in building automation system for improving building operational performance. *Energy and Buildings*, *75*, 109–118.
Yan, K., Shen, W., Mulumba, T., & Afshari, A. (2014). ARX model based fault detection and diagnosis for chillers using support vector machines. *Energy and Buildings*, *81*, 287–295.
Yu, Z., Haghighat, F., Fung, B. C. M., & Yoshino, H. (2010). A decision tree method for building energy demand modeling. *Energy and Buildings*, *42*, 1637–1646.
Yu, Z., Haghighat, F., Fung, B. C. M., Morofsky, E., & Yoshino, H. (2011). A methodology for identifying and improving occupant behavior in residential buildings. *Energy*, *36*, 6596–6608.
Yu, Z., Fung, B. C. M., Haghighat, F., Yoshino, H., & Morofsky, E. (2011). A systematic procedure to study the influence of occupant behavior on building energy consumption. *Energy and Buildings*, *43*, 1409–1417.
Yu, Z., Haghighat, F., Fung, B. C. M., & Zhou, L. (2012). A novel methodology for knowledge discovery through mining associations between building operational data. *Energy and Buildings*, *47*, 430–440.
Yu, Z., Fung, B. C. M., & Haghighat, F. (2013). Extracting knowledge from building-related data - A data mining framework. *Building Simulation*, *6*, 207–222.
This paper established a four-component framework consisting of DM techniques/algorithms (including classification analysis, cluster analysis, and ARM), potential applications, input and output. A step-by-step data analysis process that starts from problem definition to knowledge discovery was also proposed. It demonstrates the feasibility of applying the process and framework in the building field.

Zhao, H. X., & Magoulès, F. (2011). New parallel support vector regression for predicting building energy consumption. In *Proceedings of 2011 IEEE symposium on computational intelligence in multicriteria decision-making*.
Zhao, H.-x., & Magoulès, F. (2012). A review on the prediction of building energy consumption. *Renewable and Sustainable Energy Reviews*, *16*, 3586–3592.
Zhao, J., Lasternas, B., Lam, K. P., Yun, R., & Loftness, V. (2014). Occupant behavior and schedule modeling for building energy simulation through office appliance power consumption data mining. *Energy and Buildings*, *82*, 341–355.