Mining Big Building Operational Data for Building Cooling Load Prediction and Energy Efficiency Improvement

Fu Xiao*, Shengwei Wang
Department of Building Services Engineering
The Hong Kong Polytechnic University
Hong Kong, China
* linda.xiao@polyu.edu.hk

Cheng Fan
Department of Construction Management and Real Estate,
Shenzhen University
Shenzhen, China

***Abstract*—This paper aims to explore the potential application of advanced DM techniques for effective utilization of big building operational data. Case studies of mining the operational data of an institutional building for cooling load prediction and operation performance improvement is presented. Deep learning-based prediction techniques, decision tree and association rule mining are adopted to analyze the operational data. The results show that useful knowledge can be extracted for forecasting 24-hour ahead building cooling load profiles, identifying typical building operation patterns and spotting energy conservation opportunities.**

***Keywords*—Big building operational data; Building energy efficiency; Building cooling load; Deep learning; Data mining.**

I. INTRODUCTION

The building sector has become one of the largest energy consumers worldwide[1]. In Hong Kong, buildings are responsible for 92% of electricity consumption in 2014[2]. Building energy efficiency has become a global urgent issue and attracted great efforts. To improve the building operational performance, Building Automation Systems(BASs) are usually installed in modern buildings, which facilitate the real-time monitoring, control and energy management. Massive amounts of building operational data are collected and stored in BAS. However, the current utilization of big building operational data is far from being effective due to the lack of suitable methods or tools for analyzing the data. Data mining(DM) is a promising solution having excellent ability in extracting useful knowledge from massive data sets[3, 4].

This paper presents two case studies on (1) using deep learning to predict 24-hour ahead building cooling load profiles and (2) extracting useful knowledge from massive building operational data using typical DM techniques. The method is developed based on the generic data analytic framework proposed in our previous study[5]. The methods are applied to analyze the data retrieved from one building in the Hong Kong Polytechnic University.

II. RESEARCH METHODOLOGY

The generic DM-based analytic framework[5] includes four phases i.e., data exploration, data partitioning, knowledge discovery and post-mining. The following sections present the details of methods used at different phases.

A. Deep learning-based prediction techniques

Deep learning is a technique allowing computational models with multiple processing layers to learn representations of data with multiple levels of abstractions[6]. This study investigates the potential of supervised and unsupervised deep learning in predicting building cooling load. Unsupervised deep learning is applied for extracting features as model inputs. Supervised deep learning is applied for developing prediction models. The performance is validated through comparisons with advanced predictive techniques, e.g., gradient boosting machines(GBM), support vector regression(SVR), extreme gradient boosting trees(XGB).

B. Data partitioning

Data partitioning refers to the process of dividing the entire data table into several subsets, each containing a number of rows. Two data mining techniques are suitable for this task. The first is clustering analysis, which partitions the massive data sets based on data similarities. The other is decision tree, which divides the data according to certain splitting criteria. This study selected the decision tree method due to its high interpretability.

C. Knowledge discovery

One of the main approach for knowledge discovery is to investigate the relationships between different variables using association rule mining. Traditional association rule mining method has heavy burden in manually discovering potentially useful rules from a large number of association rules obtained and determining the discretization methods for numeric variables. This study adopts the quantitative association rule mining algorithm QuantMiner[7] to mine both numeric and categorical variables directly.

III. CASE STUDY

A. Description of building, system and data

The data to be analyzed are retrieved from a campus building in the Hong Kong Polytechnic University. The gross

floor area is around 11,000m^2 and 8,500m^2. The chiller plant contains 4 water-cooled chillers (denoted as CH-1 to CH-4) and 4 cooling towers (denoted as CT-1 to CT-4). Chillers are connected in parallel and the chilled water is distributed using 6 primary chilled water pumps (denoted as PCHWP-1 to 6) and 6 secondary chilled water pumps (denoted as SCHWP-1 to 6). The condenser water is circulated between chillers and 4 cooling towers using 6 VSD pumps (denoted as CDWP-1 to 6).One-year building operational data in 2015 are collected with a collection interval of 30-minute. The variables included in this dataset contains five time variables (i.e., Month, Day, Hour, Minute and Day type), the outdoor temperature, the outdoor relative humidity, the supply and return chilled water temperature and the flow rate of the chilled water temperature. The building cooling load is calculated based on the latter three variables. In total, the dataset contains 15,792 observations.

B. Building cooling load prediction using deep learning method

Building cooling load is heavily influenced by two kinds of factors, i.e., building occupancy and outdoor condition. This study considered the occupancy influence using time variables, as the occupancy schedule for a specific functional building is usually fixed and correlated with time. Therefore, the BASIC feature set contains all the five time variables (i.e., Month, Day, Hour, Minute and Day type), the outdoor temperature and the outdoor relative humidity at time T. These seven features are taken as model inputs to predict building cooling load at time T. Compared to the BASIC feature set, additional information of building cooling load, outdoor temperature and RH during the past 24-hour are added for analysis, either in their raw form or after feature extraction. If without feature extraction, each time series of building cooling load, outdoor temperature and outdoor RH during the past 24-hour will result in 48 more variables (due to a collection interval of 30-minute). The resulting feature set therefore contains 151 (i.e., 144+7) variables and is denoted as the RAW feature dataset. Another feature set is constructed using unsupervised deep learning. A deep auto-encoder model is developed for each of the three time series. An optimization process is performed to determine the optimal model architecture. The model uses a tanh activation function, i.e., tanh(z)=(e^z-e^(-z))/(e^z+e^(-z)). The resulting feature set is denoted as DAE. The entire dataset is divided into training, validation and testing data with proportions of 70%, 15% and 15% respectively. The model parameters of each algorithm are optimized through cross-validation and parameter grid search.

C. Data partitioning using decision tree method

A decision tree model is constructed using the cooling load data as shown in Fig. 1. Starting from Node 1, the model first picks the Hour as the splitting variable and the splitting criterion is {0, 1, 2, 3, 4, 5, 6, 7, 23} and {8 to 22}, which corresponds to the non-peak and peak hours. The lectures normally start at 8:30am and end at 9:30pm. Node 3 divides the data based on the Day type and the partitioning is made based on {Monday to Saturday} and {Sunday}. It should be noted at many classes for part-time students and academic activites are scheduled in this campus building on Saturdays and therefore, the cooling load on Saturdays is very similar to that on a typical weekday. Node 5 selects Month as the splitting variable and the two splitting sets are {1, 2, 3, 4, 12} and {5, 6, 7, 8, 9, 10, 11}. The first set corresponds to the cooler and less humid seasons while the second refers to the hotter and more humid seasons in Hong Kong. The splitting criteria generated at Nodes 1, 3 and 5 are further used together to partition the data in a more detailed manner. As a result, the entire data sets are partitioned into 8 groups, as shown in Table 1.

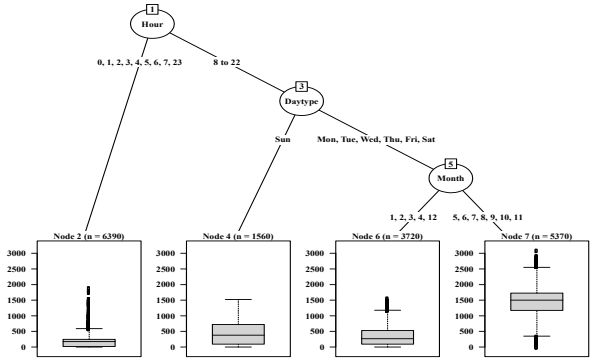

Fig. 1. Decision tree model for building cooling load

TABLE I. DETAILS ON EIGHT DATA GROUPS

Groups	Month	Day type	Hour
1	{1,2,3,4,12}	{Monday to Saturday}	{0,1,2,3,4,5,6,7,23}
2	{1,2,3,4,12}	{Monday to Saturday}	{8 to 22}
3	{1,2,3,4,12}	{Sunday}	{0,1,2,3,4,5,6,7,23}
4	{1,2,3,4,12}	{Sunday}	{8 to 22}
5	{5 to 11}	{Monday to Saturday}	{0,1,2,3,4,5,6,7,23}
6	{5 to 11}	{Monday to Saturday}	{8 to 22}
7	{5 to 11}	{Sunday}	{0,1,2,3,4,5,6,7,23}
8	{5 to 11}	{Sunday}	{8 to 22}

D. Knowledge discovery using quantitative association rule mining

The QuantMiner algorithm is adopted to discover the associations in each data group separately. Both sides of the rule, i.e. the antecedent and the consequence, are constrained to have one variable only. The genetic algorithm parameters are set as follows: 250 as the population size, 100 as iteration number, 50% as crossover rate and 40% as mutation rate[7]. The support and confidence thresholds are set as 5% and 90% respectively. In general, the confidence threshold should be set no less than 0.8 to ensure the quality of association rules.

IV. RESULTS AND DISCUSSIONS

A. Prediction of building energy performance

Table 2 summarizes the prediction performance using three metrics, i.e., the mean absolute error(MAE), the root mean squared error(RMSE) and the coefficient of variation of the root mean squared error(CV-RMSE). In terms of prediction techniques, In general, XGB method has a performance edge over the GBM, SVR and the DNN methods. The best prediction performance is achieved when XGB models are developed using the DAE feature set and the resulting CV-RMSE is 17.8%.

TABLE II. 24-HOUR AHEAD PREDICTION ACCURACY ON TESTING DATA

Method	Metrics	BASIC	RAW	DAE
GBM	RMSE	136.8	146.3	117.8
	CV-RMSE	22.8%	24.4%	19.7%
	MAE	94.2	102.5	83.9
SVR	RMSE	143.5	137.8	113.8
	CV-RMSE	24.0%	23.0%	19.0%
	MAE	109.1	98.5	85.4
XGB	RMSE	129.0	116.6	**106.5**
	CV-RMSE	21.5%	19.5%	**17.8%**
	MAE	85.8	82.1	**71.6**
DNN	RMSE	175.7	131.4	123.5
	CV-RMSE	29.3%	21.9%	20.9%
	MAE	111.9	90.2	100.5

B. Identification of energy conservation opportunities

Two examples rules presented in Table 3 indicate that when one chiller is switched on, its chilled water and condensing one chiller is switched on, its chilled water and condensing water flow rates become nearly constant. By checking the actual motor speed of PCHWP and CDWP, it is found that the motor frequency was maintained at 40Hz during operation, which means the energy saving potential of variable speed operation was not realized. This finding helps to spot the energy conservation opportunity in operations. Optimal control strategy should be developed to optimize the pressure set-point for pump speed control according to the actual cooling demand.

TABLE III. QUANTITATIVE ASSOCIATIONS FOR SYSTEM PERFORMANCE EVALUATION

No.	Antecedent	Consequent	Support (%)	Confidence (%)	Lift	Data group
1	CH1_Status = On	CH1_CHW_Flow in [47.1, 51.3]	60.1	99.5	1.7	6
2	CH2_Status = On	CH2_CDW_Flow in [46.8, 51.9]	26.1	99.7	3.8	6
3	CH_1 = Off	Performance = Good	52.5	81.1	1.3	6

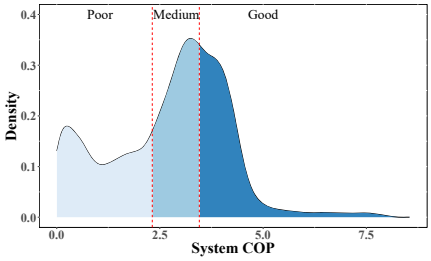

Fig. 2. Density plot of system COP

C. Evaluation on HVAC operational performance

The HVAC operational performance can be evaluated using the system coefficient of performance (COP), which equals the ratio between the building cooling load and the power consumption of the chiller plant. The equal-width binning method is applied to discretize the system COP into 3 classes, Poor, Medium and Good. Fig. 2 presents system COP data discretization results with two cutting points 2.3 and 3.5 (red vertical dashed lines). Rule No. 3 in Table 3 presents an interesting association between system components and system performance. It states that if CH-1 is switched off, then the system performance is Good. This rule initiates a hypothesis that CH-1 is less energy-efficient compared to the other 2 water-cooled chillers (note that CH-4 is not in operation in Group 6). To further investigate on this hypothesis, Fig. 3 is drawn to compare the system COP when CH-1, CH-2 and CH-3 are in operation alone. It is evident that CH-2 and CH-3 perform better than CH-1. Maintenance on CH-1 is recommended, such as cleaning the evaporator and condenser.

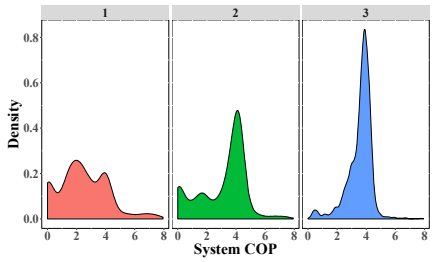

Fig. 3. Density plot of system COP when CH-1 to 3 is in operation

V. CONCLUSIONS

This paper presents the potential application of data mining in predicting building cooling load and discovering useful knowledge from massive building operational data. The method is developed from the generic data mining-based analytic framework proposed in our previous work. Results show that the proposed method can achieve accurate and reliable predictions on 24-hour ahead building cooling load profiles and extract valuable knowledge from building operational data for enhancing building energy efficiency. Further study will focus on exploiting the more recent advanced DM techniques, developing specific methods for mining big building operational data and discovering applications in building energy management.

ACKNOWLEDGMENT

The authors gratefully acknowledge the support of this research by the Research Grants Council (RGC) of the Hong Kong SAR (152181/14E).

REFERENCES

[1] Urge-Vorsatz D, Cabeza LF, Serrano S, Barreneche C, Petrichenko K. Heating and cooling energy trends and drivers in buildings. Renew Sust Energ Rev 2015; 41: 85-98.

[2] Fan C, Xiao F. Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. Appl Energ 2014;127:1-10.

[3] Liao SH, Chu PH, Hsiao PY. Data mining techniques and applications-A decade review from 2000 to 2011. Expert Syst Appl 2012;39:11303-11.

[4] Witten IH, Frank E, Hall MA. Data mining: Practical machine learning tools and techniques. 3rd ed. Massachusetts: Morgan Kaufmann; 2011.

[5] Fan C, Xiao F, Yan CC. A framework for knowledge discovery in massive building automation data and its application in building diagnostics. Automat Constr 2015;50:81-90.

[6] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015; 321: 436-44.

[7] Salleb-Aouissi A, Vrain C, Nortet C. QuantMiner: A genetic algorithm for mining quantitative association rules. IJCAI 2007：1035-40.