A Data Mining Approach to Discover Critical Events for Event-Driven Optimization in Building Air Conditioning Systems

Wang, Junqi; Zhou, Pei; Huang, Gongsheng; Wang, Wei

Published in:
Energy Procedia

Published: 01/12/2017

Document Version:
Final Published version, also known as Publisher's PDF, Publisher's Final version or Version of Record

Published version (DOI):
10.1016/j.egypro.2017.12.680

Publication details:
Wang, J., Zhou, P., Huang, G., & Wang, W. (2017). A Data Mining Approach to Discover Critical Events for Event-Driven Optimization in Building Air Conditioning Systems. *Energy Procedia*, *143*, 251-257. https://doi.org/10.1016/j.egypro.2017.12.680

Available online at www.sciencedirect.com

ScienceDirect

Energy Procedia 143 (2017) 251–257

Energy Procedia

www.elsevier.com/locate/procedia

World Engineers Summit – Applied Energy Symposium & Forum: Low Carbon Cities & Urban Energy Joint Conference, WES-CUE 2017, 19–21 July 2017, Singapore

A Data Mining Approach to Discover Critical Events for Event-Driven Optimization in Building Air Conditioning Systems

Junqi Wang [a,*], Pei Zhou [a], Gongsheng Huang [a], Wei Wang [a]

[a]Department of Architecture and Civil Engineering, City University of Hong Kong Tat Chee Avenue, Hong Kong SAR, China

Abstract

While online optimal control is regarded as an efficient tool to improve the operating efficiency of air conditioning, traditional optimal control strategies utilize the so-called time-driven optimization (TDO) scheme which triggers actions by “time”. Although it works well for simple air conditioning systems, several limitations are encountered when systems become more and more complex. Since TDO is a periodic scheme, it may not be suitable or efficient to react to stochastic operational changes. Recently, in order to solve those limitations, the event-driven optimization (EDO) scheme has been proposed, in which actions are triggered by “event”. However, previous studies only used prior knowledge to discover important events, which could only find events for general systems, and might not comprehensive because human prior knowledge is limited after all. Moreover, prior-knowledge-based method is able to discover new knowledge. Thus, this paper presents an effective data mining approach to discover the hidden knowledge in massive data set for EDO in building air conditioning systems. Results shown that data-mining-based EDO achieves a higher energy saving with reduced computation load, in comparison with the traditional TDO. Since the data mining approach can help to automate the process of finding critical events and event threshold, it also improves the practicability of EDO.

Peer-review under responsibility of the scientific committee of the World Engineers Summit – Applied Energy Symposium & Forum: Low Carbon Cities & Urban Energy Joint Conference.

Keywords: *data mining; variable importance; building efficiency; event-driven optimization; air conditioning systems;*

* Corresponding author. Tel.: +852-63589502
E-mail address: jqwang8-c@my.cityu.edu.hk

1876-6102 © 2017 The Authors. Published by Elsevier Ltd.
Peer-review under responsibility of the scientific committee of the World Engineers Summit – Applied Energy Symposium & Forum: Low Carbon Cities & Urban Energy Joint Conference.
10.1016/j.egypro.2017.12.680

1. Introduction

Event-driven optimization (EDO) has found to be effective in achieving energy efficient operation with greatly reduced computation load, which makes it applicable for online optimal control of air conditioning systems. In EDO, “event” plays an important role since it drives the timing and action of the optimization, which is a key factor affecting the energy and computation efficiencies of online optimal control.

The majority of the existing operation strategies for building services systems is developed by domain expertise or heuristic rules [1,2]. Thus, in our previous studies [3,4], prior knowledge was utilized and important events were successfully identified. However, considering human prior knowledge is limited and the system is not static (also evolving), prior understandings may be insufficient when pursuing further improvement, and additional ways are needed to discover more critical events for EDO. With the widespread penetration of building automation systems (BASs), huge amount BAS data is available. Effective understanding of these data can help to find important events. Data mining is believed to be a powerful tool to learn the hidden knowledge inside large dataset [5]. The major data mining practices in building field are as follows: finding patterns, associations, or relationships [6], building prediction models [7,8], diagnostics [9] and tuning controllers [10].

However, most of the previous attempts stop at identifying critical factors on energy performance or system faults. For building optimal control, only guidelines were provided, while the direct benefit of applying the discovered knowledge has rarely been demonstrated. Therefore, this study presents a more complete data mining practice on BAS data, which investigates the dataset, interprets the dataset, and utilizes the discovered knowledge to formulate the optimal control strategy (under the EDO framework), where the benefits obtained from the formulated strategy are directly revealed through simulation. Following contributions are made in this paper: (1) variable importance is directly output by the data mining algorithm, and users can select events based on the relative importance instead of domain knowledge; (2) Euclidean distance of decision variables is proposed to estimate the optimization reward; (3) event threshold is directly computed with the help of Euclidean distance of decision variables, which avoids the tedious trial-and-error method previously used [4].

2. Method

2.1. Brief overview of EDO and event space establishment

EDO is a new RTO framework in which optimization actions are triggered by event [3,4]. An event describes a set of *state transitions* that physically happen in a system [11,12]. In daily operation of air conditioning systems, state transitions are numerous and may come from environment (e.g. weather changes), system itself (e.g. operation mode changes) and occupants (e.g. occupancy changes). As not all of the state transitions should be used to define events in EDO, only those important state transitions that could cause a significant influence on concerned objectives will be selected. The selection criteria is developed using the notion of *optimization reward* as follows.

- If the predicted *optimization reward* associated with the state transition is “large”, then we select it;
- If the predicted *optimization reward* associated with the state transition is “small”, then we will not use it (considering that it would also consume resources like computation).

To establish the event space (the set of events), a three-step method is developed as shown in Figure 1. Step 1 is state transition identification that is used to identify possible critical state transitions. Step 2 is to define the candidate event space. Step 3 is to optimize the candidate event space by event performance analysis. Finally, the EDO design will be validated. Please note that although a state transition could be important, it may be a suitable event. Therefore, the Step 3 is necessary, which performs the analysis of event performance in terms of three key indices, i.e. energy performance, computational performance and performance score. Performance score is a index that both considers energy and computational performances (shown in eqn. (1)).

$$PS = a \times score_{EC} + b \times score_{CC} \tag{1}$$

where $score_{EC} = ES\%_{ei} / ES\%_{BC}$ and $score_{CC} = CS\%_{ei}$; a and b are the weighting factors; *ES%* is energy saving and *CS%* is computation saving; *EC* is energy consumption and *CC* is computation consumption; *BC* is the base case; ei is the event i;

In this study, "$a = 1$" and "$b = 0.5$" were used as weighting factors in eqn. (1) simply because our priority was given to the energy performance.

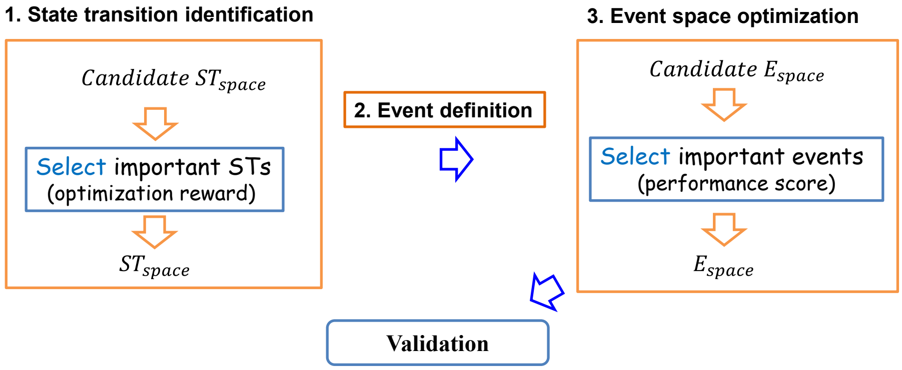

Figure 1 The three-step method to establish event space

2.2. Discovering critical events through data mining

2.2.1. Measures of optimization reward

To proceed, assume (i) there are n decision variables $V = v_1, \cdots, v_n$ adopted in the real-time optimization; (ii) there are m state variables $S = s_1, \cdots, s_m$ (from climatic conditions, system operating status and occupants' behaviours); (iii) at time k, the optimal values of these n decision variables are $V_k = v_{1,k}, \cdots, v_{n,k}$, and system state is $S_k = s_{1,k}, \cdots, s_{m,k}$; (iv) time k-1 is the previous optimization time instance; (v) SCOP is the system COP, and the *optimization objective* is to achieve the highest SCOP. Typically, we would have the following observations in system operation:

(1) if $S_k = S_{k-1}$, then $V_k = V_{k-1}$ (or $|V_k - V_{k-1}| = 0$), $SCOP_k^{V_k} = SCOP_{k-1,}^{V_{k-1}}$;

(2) if $S_k \neq S_{k-1}$, then $V_k \neq V_{k-1}$ (or $|V_k - V_{k-1}| > 0$), $SCOP_k^{V_k} > SCOP_k^{V_{k-1}}$.

The above observations are further illustrated in Figure 2. If the system state remains the same at time k-1 and k (case a of Figure 2), $V_k = V_{k-1}$ and SCOP remains the same. When the system state varies (case b of Figure 2), the previous settings V_{k-1} would not be optimal for the condition at time k. Therefore, we take the optimization. The decision variable vector is optimized to V_k, and the SCOP would be improved comparing with $SCOP_k^{V_{k-1}}$ (SCOP at time k using V_{k-1}). The performance gap between $SCOP_k^{V_k}$ and $SCOP_k^{V_{k-1}}$ is called *optimization reward* (highlighted in Figure 2). A reasonable assumption is that: the more system state deviates from the previous time, the more the decision variable vector would deviate from the previous one, and thus, the higher performance reward would be. Next, we will develop quantities to mathematically describe the deviations between adjacent decision variables and system states.

In data mining, the deviation between objects is usually called "dissimilarity" [13]. The commonly used dissimilarity measure is Euclidean distance (others are Minkowski and Mahalanobis distance), which is defined as:

$$d_k = \sqrt{(v_{1,k} - v_{1,k-1})^2 + \cdots + (v_{n,k} - v_{n,k-1})^2} \quad (2)$$

To eliminate the scale effect in dataset, data normalization can be performed and the Euclidean distance is re-written as:

$$d_{k,norm} = \sqrt{(v_{1,k,norm} - v_{1,k-1,norm})^2 + \cdots + (v_{n,k,norm} - v_{n,k-1,norm})^2} \tag{3}$$

$$v_{n,norm} = (v_n - v_{n,\min})/(v_{n,\max} - v_{n,\min}) \tag{4}$$

Simultaneously, the variations of the monitoring state variables are:

$$\Delta S_k = (\Delta s_{1,k}, \cdots, \Delta s_{m,k}), \text{ where } \Delta s_{i,k} = |s_{i,k} - s_{i,k-1}|, i = 1, \cdots, m. \tag{5}$$

The Euclidean distance can be used to estimate the optimization reward. To investigate how the state variables affect the optimization reward, we can write the Euclidean distance of decision variables as a function of state variables and their variations (eqn. (6)).

$$d_{k,norm} = f(S_k, \Delta S_k) = f(s_{1,k}, \cdots, s_{m,k}, \Delta s_{1,k}, \cdots, \Delta s_{m,k}) \tag{6}$$

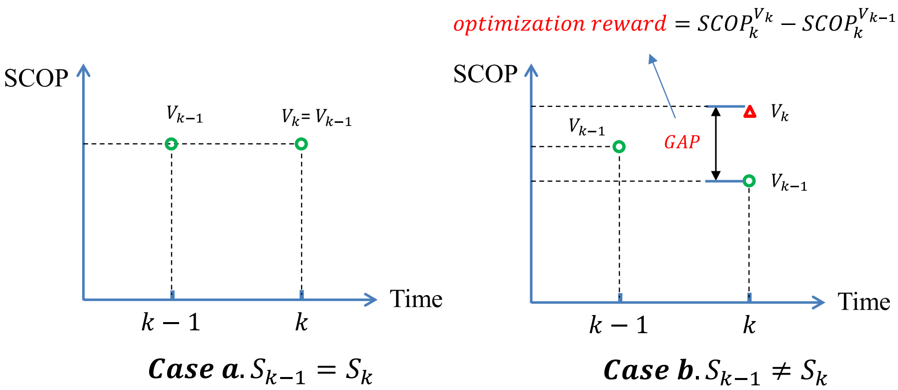

Figure 2 Illustration of optimization reward

2.2.2. Finding important variables with Random forest

To find how state variable changes would affect the optimization reward, variable importance should be studied. Thus, random forest (RF) is used in this paper, which is an ensemble learning method that constructs multiple decision trees based on random selection of data and variables. It is popular for classification and regression because it can handle sparse and correlated data. Moreover, RF has an inherent procedure to evaluate variable importance, which is perfect for the current application. The "randomForest" package (version 4.6-12) in R (version 3.3.3) [14] is directly used to implement the RF algorithm. The percentage increase of the mean square error (%IncMSE) is computed by the package, and a higher %IncMSE means a higher variable importance [15].

2.2.3. Finding the thresholds of events based on decision variable distance

For the event defined by continuous variables, thresholds are needed to quantify the variable variation (ΔS_k). Since the event should associate with the "large" optimization reward (Section 2.1), large $d_{k,norm}$ is firstly defined. Then, the corresponding state variables and variable variations are located by the index of large $d_{k,norm}$ (or $Index_{large}$). As in eqn. (7), "large" distance is defined as the distance greater than the mean of all the decision variable distance. Generally, the event threshold is a user-defined parameter, and the threshold value directly affects the event triggering. In this paper, the mean of the state variable (or state variable variation) is suggested (eqn. (10-11)) as the event threshold since the average could represent the typical value in a set of data. (Note: $|A|$ means the size of data set A)

$$d_{norm,large} := \{\forall d_{norm,j} > d_{norm,mean}, j \in \{1,2,...,|data\ set|\}\} \tag{7}$$

$$Index_{large} := \{\forall j, d_{norm,j} > d_{norm,mean}, j \in \{1,2,...,|data\ set|\}\} \tag{8}$$

$$\Delta S_{large} := \{\Delta s_j, j \in Index_{large}\},\ S_{large} := \{s_j, j \in Index_{large}\} \tag{9}$$

$$Threshold\ s_j := \{\sum_j s_j / |Index_{large}|, j \in Index_{large}\} \tag{10}$$

$$Threshold\ \Delta s_j := \{\sum_j \Delta s_j / |Index_{large}|, j \in Index_{large}\} \tag{11}$$

3. Case study

3.1. System description

The air conditioning system used in the case study is a typical all-electric system which does not contain significant thermal storage. The system is established according to a supertall office building in Hong Kong. Circulation loops of cooling water, chilled water primary side and secondary side are simulated in details, while the air distribution system is simplified by one zone. The validated models of cooling towers, chillers and pumps established in [16-18] were used. The optimal control strategy was coded in the Matlab module, and the so-called co-simulation between TRNSYS and Maltab was adopted to simulate the system operation based on real load profiles obtained from site measurement.

3.2. Data preparation

In general, the data from a BAS with TDO function can be used for data mining using the method in Section 2.2. In this paper, the data generated from the simulation platform was used as an alternative, where the same method can be adopted to identify the important variables and corresponding thresholds. The allowable highest optimization frequency 15 minutes per optimization was used for TDO, which was also used as a benchmark when evaluating the performance of EDO. The real weather and load profiles on May 13-17, 2013 were used to generate data for data mining. Then, events were defined based on the data mining results and used to formulate the EDO strategy. This EDO strategy was performed on the simulation platform using the load profile of May 20, 2013 in order to validate the effectiveness.

3.3. Results

Table 1 Variable importance (%IncMSE) from random forest (Rank 1 means the highest importance)

Variable	%IncMSE	Rank	Variable description
Δh	26.84	1	the average difference between the specific enthalpies of saturated air and bulk air
T_{apr}	21.88	2	the cooling tower approach temperature
$Freq_{ct}$	18.62	3	the fan frequency of cooling tower
PLR	15.42	4	part-load ratio
$M_{w,prm,HX}$	14.69	5	water mass flow rate of the heat exchanger at primary sided
$M_{w,sec,pump}$	11.24	6	water mass flow rate of total secondary pumps
T_{wb}	3.75	7	wet-bulb temperature
T_{db}	3.4	8	dry-bulb temperature;
$Num_{chiller}$	1.13	9	number of the operating chiller(s)

(Note: 500 trees were used and 3 variables were tried at each split in random forest.)

Table 2 Optimization performance of different events (Op. = Optimization; Ch. = Chiller.)

Op. methods	**Energy consumption (kwh)**	**Energy Saving**	**Op. time (s)**	**Computation Saving**	**Threshold**	**Performance Score**
No Op.	158197	0.00%	0	\		
15mins	142329	**10.03%**	130.6	0.00%		
PLR Change	142325	10.03%	55.35	57.62%	6%	1.288
$\mathbf{\Delta h}$	142005	10.24%	76.38	41.52%	36 kJ/kg	1.228
T_{apr}	142710	9.79%	68.7	47.40%	5.5 ºC	1.213
$Freq_{ct,fan}$	142305	10.05%	145.1	*-11.10%*	36	0.946
$\mathbf{M_{w,prm,HX}}$ **Change**	142266	10.07%	28.74	77.99%	4.7 L/s	1.394
$M_{w,sec,pump}$ Change	143277	9.43%	71.62	45.16%	18 L/s	1.166
PLR Change, $M_{w,prm,HX}$ Change & Δh	140732	**11.01%**	82.6	62.75%	6%; 4.7 L/s; 36 kJ/kg	1.411

As shown in Table 1, nine variables were evaluated and listed according to their rankings. A higher "%IncMSE" value means a higher importance of the variable. The variable importance (%IncMSE) of "T_{wb}", "T_{db}" and "$Num_{chiller}$" are lower than 5%, which are considered as little effect on the optimization reward. Thus, the three state variables are not considered in the event definition.

The other six variables were defined as events and tested in the simulation. Optimization performance of events are shown in Table 2. To avoid event duplication, only the event with highest importance was selected from the same event source. "Δh", "T_{apr}" and $Freq_{ct}$ all reflect the heat rejection processes in cooling towers [19]. Thus, only the most important one (i.e. "Δh") was chosen. Similarly, "$M_{w,prm,HX}$ Change" was used, while "$M_{w,sec,pump}$ Change" was discarded. For event threshold, based on eqn. (10) and (11), "6%" was used for "PLR change", "4.7 L/s" was used for "$M_{w,prm,HX}$ Change" and "36 kJ/kg" was used for "Δh". To validate the optimization performance of the established event space, computer simulation was performed to compare the EDO design (using the established event space) with traditional TDO. Result is presented in Table 2, last row. It shows that, by adopting the event space output from the data mining, the energy saving of EDO (11.01%) is almost 1% higher than the traditional TDO (10.03%) using the optimization frequency of 15 minutes. Meanwhile, 62.75% of computation was saved.

3.4. Discussions

The performance score (including energy and computational performances) of the discovered events are all better than the TDO benchmark, except the case of "$Freq_{ct,fan}$". "$Freq_{ct,fan}$" achieved a negative computation saving because it triggers too many optimization, which is considered as frequent event triggering. The possible reason is that the threshold selection is unsuitable, which is near the fluctuating point of the variable.

The proposed index, Euclidean distance of decision variables, is an effective estimator of the optimization reward. Using this index, significant optimization reward can be quickly estimated in the data set. Suitable event threshold can be directly computed with this Euclidean distance, which avoids the tedious trial-and-error method used in the knowledge-based method. Compared with the knowledge-based method, new events are found in the operational data, namely "$Freq_{ct,fan}$", "$M_{w,prm,pump}$ Change" and "$M_{w,sec,pump}$ Change".

4. Conclusion

The selections of events and event thresholds are critical for the performance of EDO. Instead of using prior knowledge, this paper has effectively explored the data mining technique for EDO in the field of building optimal control. Optimization reward is formally introduced in this paper, which is a key notion to facilitate the selection of state transition. The variable importance output by the random forest algorithm is utilized to define and select events, which enables the quantitative evaluation of contributions of different state variables to the optimization reward. A new index, Euclidean distance of decision variables, is proposed to estimate the optimization reward, which offers a simple way for event threshold selection and avoids the trial-and-error methods previously used in [4, 20]. Compared with the traditional TDO, the EDO based on data mining achieves a higher energy saving with reduced computation load. It has been found that data mining can greatly improve the practicability of EDO in terms of finding important events and suitable event thresholds for practical applications. The proposed method can customize the event space for a specific system because the operational data is used. The proposed method is easy to use and general as long as enough BAS data is available. The method can also can be applied in existing buildings to improve the operating efficiency by identifying important state transitions and formulating more efficient operation strategies.

Acknowledgements

This research was supported by a grant of Fundamental and Application Research Special Funding of Guang Dong Province 2015 (Project No. 2015A030313814).

References

[1] Fan C, Xiao F, Madsen H, Wang D. (2015) "Temporal knowledge discovery in big BAS data for building energy management". *Energy Build* 2015;109:75-89.
[2] ASHRAE. (2015) "CHAPTER 42 SUPERVISORY CONTROL STRATEGIES AND OPTIMIZATION". *ASHRAE Handbook - HVAC Applications.*, Atlanta: USA: ASHRAE Inc. (2015).
[3] Wang J, Huang G, Sun Y. (2016) "Optimal Control of Complex HVAC Systems: Event-driven or Time-driven Optimization?". *CLIMA 2016* (2016).
[4] Wang J, Huang G, Sun Y, Liu X. (2016) "Event-driven optimization of complex HVAC systems". *Energy Build* (2016);133:79-87.
[5] Yu Z, Fung BC, Haghighat F. (2013) "Extracting knowledge from building-related data—A data mining framework". *Build Simu* (2013);6:207-22.
[6] Fan C, Xiao F, Yan C. (2015) "A framework for knowledge discovery in massive building automation data and its application in building diagnostics". *Autom Constr* (2015);50:81-90.
[7] Yu Z, Haghighat F, Fung BC, Yoshino H. (2010) "A decision tree method for building energy demand modeling". *Energy Build* (2010);42:1637-46.
[8] Zhao H, Magoulès F. (2012) "A review on the prediction of building energy consumption". *Renewable and Sustainable Energy Reviews* (2012);16:3586-92.
[9] Xiao F, Fan C. (2014) "Data mining in building automation system for improving building operational performance". *Energy Build* (2014);75:109-18.
[10] Hussain S, Gabbar HA, Bondarenko D, Musharavati F, Pokharel S. (2014) "Comfort-based fuzzy control optimization for energy conservation in HVAC systems". *Control Eng Pract* (2014);32:172-82.
[11] Cassandras CG, Lafortune S. (2009) "Introduction to discrete event systems." *Springer Science & Business Media*, (2009).
[12] Xia L, Jia Q, Cao X. (2014) "A tutorial on event-based optimization—a new optimization framework". *Discrete Event Dynamic Systems* (2014);24:103-32.
[13] Han J, Pei J, Kamber M. (2011) "Data mining: concepts and techniques." *Elsevier*, (2011).
[14] R Project for Statistical Computing (version 3.3.3) , 2017 (Available online: https://www.r-project.org/).
[15] Yu F, Ho W, Chan K, Sit R. (2017) "Critique of operating variables importance on chiller energy performance using random forest". *Energy Build* (2017);139:653-64.
[16] Ma Z. (2008) "Online supervisory and optimal control of complex building central chilling systems." Doctoral dissertation, *The Hong Kong Polytechnic University* (2008).
[17] Ma Z, Wang S. (2009) "An optimal control strategy for complex building central chilled water systems for practical and real-time applications". *Build Environ* (2009), 44(6): 1188-1198.
[18] Sun Y, Huang G, Li Z, Wang S. "Multiplexed optimization for complex air conditioning systems". *Build Environ* (2013), 65: 99-108.
[19] Chang C, Shieh S, Jang S, Wu C, Tsou Y. (2015) "Energy conservation improvement and ON–OFF switch times reduction for an existing VFD-fan-based cooling tower". *Appl Energy* (2015);154:491-9.
[20] Wang J, Huang G, Zhou P. (2017) "Event-driven optimal control of complex HVAC systems based on COP·mins", *Energy Procedia* (2017), 105:2373-2378