Energy and Buildings 133 (2016) 37–45

Contents lists available at ScienceDirect

Energy and Buildings

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

A decision tree based data-driven diagnostic strategy for air handling units

Rui Yan [a], Zhenjun Ma [a,*], Yang Zhao [b], Georgios Kokogiannakis [a]

[a] Sustainable Buildings Research Centre (SBRC), University of Wollongong, NSW, 2522, Australia
[b] Institute of Refrigeration and Cryogenics, Zhejiang University, China

ARTICLE INFO

Article history:
Received 11 May 2016
Received in revised form 10 August 2016
Accepted 20 September 2016
Available online 21 September 2016

Keywords:
Decision tree
CART
Feature selection
Air handling unit
Fault diagnosis
Interpretability

ABSTRACT

Data-driven methods for fault detection and diagnosis of air handling units (AHUs) have attracted wide attention as they do not require high-level expert knowledge of the system of concern. This paper presents a decision tree based data-driven diagnostic strategy for AHUs, in which classification and regression tree (CART) algorithm is used for decision tree induction. A great advantage of the decision tree is that it can be understood and interpreted and therefore its reliability in fault diagnosis can be validated by both testing data and expert knowledge. A steady-state detector and a regression model are incorporated into the strategy to increase the interpretability of the diagnostic strategy developed. The proposed strategy is validated using the data from ASHRAE 1312-RP. It is shown that this strategy can achieve a good diagnostic performance with an average F-measure of 0.97. The interpretation of the diagnostic decision tree using expert knowledge showed that some diagnostic rules generated in the decision tree comply with expert knowledge. Nevertheless, the interpretation also indicated that some diagnostic rules generated are not reliable and some of them are only valid under certain operating conditions, which indirectly demonstrated the importance of the interpretability of fault diagnostic models developed using data-driven methods.

1. Introduction

The operation of building Heating, Ventilation, and Air-Conditioning (HVAC) systems is vulnerable to various faults, and the occurrence of any fault could lead to increased energy consumption or indoor thermal discomfort [1,2]. Successful detection and isolation of any faults in HVAC systems in a timely manner can improve building energy efficiency and reduce carbon footprint.

Over the last two decades, considerable efforts have been made in the development of fault detection and diagnosis (FDD) strategies for building HVAC systems [3–5]. Fault detection is a process that determines whether or not an HVAC system is operating in a healthy condition while fault diagnosis aims to identify the causes of the fault. Compared to fault detection, fault diagnosis is more challenging and complicated. The existing FDD methods can be generally categorised into model-based methods, rule-based methods and data-driven methods [6]. Data-driven methods have attracted wide attention as they do not require high-level expert knowledge of the system and the computational costs are generally manageable [4].

Air handling unit (AHU) is one of the important components in HVAC systems and a number of studies have been focused on detection and diagnosis of various faults in AHUs. For instance, a rule-based FDD strategy for AHU temperature sensors was developed by Yang et al. [7], in which a set of if-then rules were formulated based on expert knowledge. The performance evaluation using the data from a green building and a small scale AHU simulator showed that this strategy is capable of isolating AHU sensor faults under different operating modes. A FDD strategy based on Bayesian Belief Network (BBN) for AHUs was presented by Zhao et al. [8,9], in which a probabilistic graphical model was used to represent the relationships of probabilistic dependencies within different variables. It was shown that this method can successfully isolate common faults occurred in AHUs. A drawback of this method is that it requires high-level expert knowledge, especially in determining the probability parameters. A sensor fault detection strategy for AHUs using cluster analysis was developed by Yan et al. [5], in which the clustering algorithm Ordering Points to Identify the Clustering Structure (OPTICS) was used to identify the spatial separated data groups which possibly indicate the occurrence of sensor faults. Wang et al. [10] described a hybrid FDD method

* Corresponding author.
E-mail addresses: mzjxjtu@163.com, zhenjun@uow.edu.au (Z. Ma).

http://dx.doi.org/10.1016/j.enbuild.2016.09.039
0378-7788/

Nomenclature

Symbols

a_1-a_3	Coefficients
α	Complexity parameter
c	Total number of class
F	Air flow rate (m^3/min)
FN	False negative
FP	False positive
I	Impurity measure
N	Fan speed control signal (%)
n	Number of observations
P	Power (W) or the proportion of the observations
p	Probability of the observations in a node
R	Misclassification rate
r	Node misclassification rate
s	Split
T	Temperature (°C)
t	Node for splitting
TP	True positive
U	Valve control signal (%)

Subscripts

cc	Cooling coil
L	Left
ma	Mixed air
oa	Outdoor air
R	Right
rf	Return fan
sa	Supply air
sf	Supply fan

which combines physical models and expert rules for variable air volume AHUs. To increase the model accuracy, genetic algorithms were used to determine the parameters of the model.

Data-driven based fault diagnosis methods were primarily developed based on pattern classification techniques [11]. The data-driven FDD methods for AHUs can be generally categorised into statistic based methods [12–14] and machine learning based methods [6,15–17]. Principal Component Analysis (PCA) and Fisher Discriminant Analysis (FDA) are two common statistic based methods. PCA is a multivariate analysis method which transforms the data set into a new set of uncorrelated variables so that the first few principal components retain the most of the variations [18]. Wang and Xiao [12], for example, presented a PCA-based AHU sensor FDD method, in which AHU sensor faults were detected by comparing the PCA model prediction error with a pre-defined threshold while fault diagnosis was achieved by using Q-contribution plot. It was shown that PCA is capable of extracting major information from the data with high dimensions and then presenting it in lower dimensions. FDA is a classifier based on linear dimensionality reduction techniques and is optimal in terms of maximising the separations among different classes. Du and Jin [14] used PCA for fault detection and FDA for fault diagnosis of AHU sensor faults. In a two-step AHU sensor fault detection and isolation strategy developed by Padilla and Choinière [19], PCA was used in the first step for fault detection while active functional testing was used in the second step for fault isolation. Among the machine learning-based methods, the use of artificial neural networks (ANNs) for AHU FDD has been extensively studied [6,20–24]. An ANN-based AHU FDD strategy was developed by Lee et al. [22]. Seven user-defined residual variables between the predicted and measured values were fed into a two-layer feed-forward ANN trained by back-propagation with both normal and faulty data. Eight typical AHU sensor and mechanical faults were considered. The results showed that this ANN-based FDD method can identify the root cause of AHU faults. An AHU FDD strategy using ANN and wavelet analysis was developed by Fan et al. [20]. Back-propagation neural network (BPNN) was used for AHU fault detection while Elman neural network coupled with wavelet analysis was used for fault diagnosis. Support vector machine (SVM) is another commonly used machine learning method for AHU FDD. Liang and Du [16] proposed a diagnostic strategy for AHUs based on support vector machine (SVM) method. This strategy can isolate different fault classes through finding the boundary that can maximize the margins between the classes. Four physical models were formulated and the residuals between the predicted and actual values were used as the input variables of the SVM classifier. It was found that the diagnostic accuracy of the SVM-based method mainly relies on the selection of the SVM kernel function. Mulumba et al. [17] described a model-based FDD strategy with SVM classification and an autoregressive model with exogenous variables (ARX) for AHU FDD. The supply air humidity was selected as a dependent variable while the supply air temperature and mixed air temperature were selected as the exogenous variables for the ARX model using ReliefF-based feature selection. The performance of the SVM-based method was compared with the classifiers such as ANN, random forest, and Naïve Bayes. It was shown that the SVM-based method outperformed the other methods in terms of the diagnostic accuracy. Wall et al. [25] described an AHU FDD method using a dynamic Bayesian network (DBN). DBN is similar to BBN but with an extra dimension of time. In contrast to the BBN-based method proposed in [8] in which the BBN parameters were estimated by expert knowledge, the parameters in DBN were obtained using both normal and faulty training data.

Although data-driven FDD methods are promising when validated using the testing data, a key drawback of this approach is that most data-driven methods (e.g. ANN and SVM-based) were developed based on black-box models, which means that it is almost impossible to understand how faults are isolated. As the actual fault diagnostic accuracy of data-driven methods heavily relies on the quality of training data used, the factors such as lacking of unique data patterns and errors in the training data may result in invalid classification. In many cases, only a limited amount of training data is available, which usually cannot cover the full operation range of the system of concern. The model developed may therefore only valid under certain operating conditions. Without interpretability, it is difficult to know whether or not the model used is reliable and under what operating conditions the model is reliable.

This paper presents a decision tree based data-driven method for fault diagnosis of AHUs. Decision tree, as a well-known classifier, has been applied in prediction of building energy usage with satisfied accuracy [26,27]. Compared to other data-driven diagnostic strategies, the proposed strategy is interpretable by using the decision tree which can generate a set of if-then rules. The interpretability of the proposed strategy can be helpful in understanding the diagnostic strategy for isolating different AHU faults. The proposed strategy can also automatically perform feature selection, which often requires considerable efforts to analyze and define the key features in data-driven FDD strategies [28].

2. Outline of the diagnostic strategy

The outline of the proposed AHU diagnostic strategy is illustrated in Fig. 1. The overall strategy consists of three steps including 1) data preparation; 2) decision tree induction and evaluation and; 3) decision tree interpretation.

The fault-free data is first used to determine the coefficients of a regression model, as shown in Eq. (1) [8]. This regression model is used to generate a residual feature in order to reduce the