Energy and Buildings 127 (2016) 540–551

Contents lists available at ScienceDirect

Energy and Buildings

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

Fault detection and diagnosis for building cooling system with a tree-structured learning method☆

Dan Li[a], Yuxun Zhou[b], Guoqiang Hu[a,*], Costas J. Spanos[b]

[a] School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore
[b] Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720, United States

ARTICLE INFO

Article history:
Received 6 January 2016
Received in revised form 4 June 2016
Accepted 6 June 2016
Available online 8 June 2016

Keywords:
Fault detection and diagnosis (FDD)
Building cooling system
Data-driven method
Pattern classification
Machine learning method

ABSTRACT

In order to save energy and improve the performance of building environment regulation, there is an increasing need for fault detection and diagnosis (FDD). This paper investigates the effectiveness of tree-structured learning method for FDD of building cooling system. Researchers have been tackling building FDD task with a wide variety of techniques, such as analytical model-based, signal-based and knowledge-based methods. Recently data-driven method has shown its advantage in dealing with complex systems with random penetrations. Existing work on data-driven FDD merely formulates the task as a pure fault type classification problem, whereas fault severity levels and their inter-dependence have long been ignored. We propose a novel data-driven strategy that adopts structured labeling to include the dependence information and describe the severity levels in a large margin learning framework. A Tree-structured Fault Dependence Kernel (TFDK) method is derived and a corresponding on-line learning algorithm is developed for streaming data. As an improvement of traditional classification methods (e.g. SVM), TFDK encodes tree-structured fault dependence in its feature mapping, and takes regularized misclassification cost as learning objective. Following the ASHRAE Research Project 1043 (RP-1043), the strategy is applied to the FDD of a 90-ton centrifugal water-cooled chiller. Experimental results show that compared to previous data-driven methods, TFDK can greatly improve the FDD performance as well as recognize the fault severity levels with high accuracy.

1. Introduction

Building energy consumption contributes to more than 40% of the total energy usage worldwide [1,2]. Almost 32% of the total energy in industrialized countries is consumed by heating, ventilation, and air-conditioning (HVAC) systems [3]. The newly published ASHRAE Handbook has put special emphasis on automated fault detection and diagnosis (FDD) for smart building systems. In particular, the new standard highlights the necessity of maintaining the whole building system in good working conditions through FDD techniques as well as the significance of saving energy and improving occupancy comfort level and building safety level via automated FDD system [4]. Therefore, there is an increasing need for studying automated fault identification in buildings aiming at saving energy and offering more comfortable and safe dwelling environment [5,6]. In the past decades, researchers have been sparing no efforts to develop algorithms and strategies that could detect and diagnose HVAC faults to prevent unnecessary economic losses and maintain the system's working efficiency [7,8].

In the literature, miscellaneous FDD methods have been proposed, mainly including three techniques and their combinations, such as analytical model-based, signal-based and knowledge-based methods [9–13]. The model-based method relies on explicit description of the system. Despite significant theoretical advancement made in this direction, few of the solutions can be directly inserted to the Building Management System (BMS) to conduct real-time monitoring [10,12]. The signal-based FDD method investigates the correlation between faults and system output signals, and improved performance can be achieved by adding the signal pattern of healthy status as a priori [10,12]. The knowledge-based FDD method discovers the underlying knowledge and system features that represent the information redundancy among the system's variables through learning from empirical data. Due to

☆ This research is funded by the Republic of Singapore's National Research Foundation under its Campus for Research Excellence and Technological Enterprise (CREATE) program through a grant to the Berkeley Education Alliance for Research in Singapore (BEARS) for the Singapore-Berkeley Building Efficiency and Sustainability in the Tropics (SinBerBEST) Program. BEARS has been established by the University of California, Berkeley as a center for intellectual excellence in research and education in Singapore.

* Corresponding author.

http://dx.doi.org/10.1016/j.enbuild.2016.06.017
0378-7788/

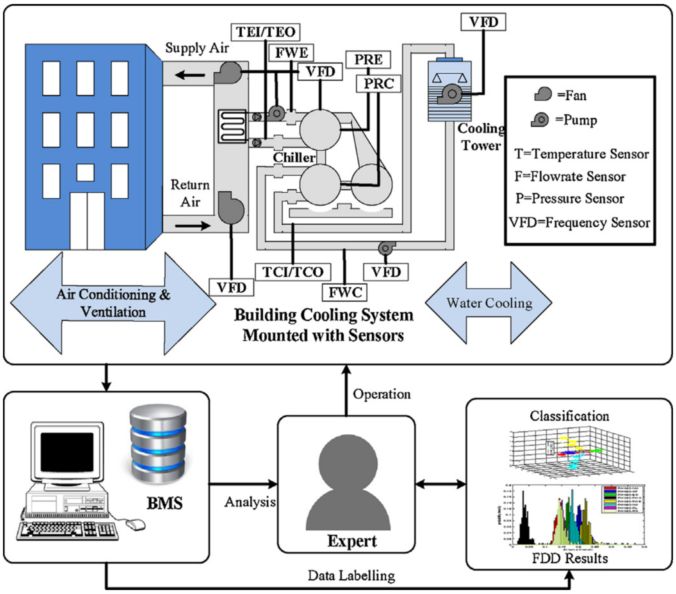

Fig. 1. Data-driven building FDD scheme, including deployed sensor network, data base management, and a decision support system.

this fact, the knowledge-based method is commonly referred to as data-driven method [10,13]. The empirical data, which records outside environmental factors, internal loads, and mechanical system working conditions, is collected through sensor network and stored in the BMS [14,15]. Experts and researchers analyze the empirical data and feedback to building operators if any fault is found. A common data-driven FDD system for smart buildings is depicted in Fig. 1, including deployed sensor network, data base management, and a decision support system.

Recently, a wide range of statistical and machine learning techniques have been explored as data-driven methods in the building FDD field, including Principal Component Analysis (PCA) [16–18], Statistical Process Control (SPC) [19–21], Multivariate Regression Models [22], Bayes Classifier [23–25], Neural Networks (NN) [26–28], Fisher Discriminant Analysis (FDA) [29], Gaussian Mixture Model [30], Support Vector Data Description (SVDD) [31,32], and Support Vector Machines (SVM) [33–37]. Among these approaches, PCA and SPC are unsupervised methods that do not require expert knowledge for fault labeling, but others like NN and FDA are supervised multi-class classification methods that depend on the availability of labeled training data. Once the hypothesis/model is fitted from the training phase, new measurements will be tested by the classifiers and be assigned to corresponding categories (normal or faulty) automatically. Notwithstanding existing work on data-driven FDD has shown promising results in both detection accuracy and efficiency, two important issues, namely fault interdependence and severity levels, are often ignored or over-simplified with homogeneity assumptions [13,38,39].

First of all, although it is quite intuitive to build fault dependence by analysing the connections and structures of each component of HVAC system, this prior knowledge is rarely considered in current data-driven FDD literature. For example, Zhao proposed a chiller fault detection method based on Support Vector Data Description (SVDD), which is a one-class classification technique describing the support of data distribution [31]. By training SVDD models for each fault type, they extended similar idea to a chiller fault diagnosis strategy in [32]. Noticing that training a one-class classification model for each specific fault type is computationally costly, an alternative method is to formulate the FDD issue directly as a multi-class classification problem. To list a few, Du proposed to utilize Fisher Discriminant Analysis (FDA) and Principal Component Analysis (PCA) to diagnose multiple sensor faults in AHU [29]. Keigo employed semi-supervised FDA to detect building energy faults, and adopted Decision Boundary Analysis (DBA) to discover the hidden relationship between the extracted features and the corresponding faults [40]. However, all of the aforementioned work is restricted to modeling each type of fault separately with single (flat) class labels and ignores valuable prior information on fault dependence, which could otherwise be exploited (fused) to improve the detection performance of the machine learning method [41]. Moreover, when dealing with complex building systems, the number of fault types (classes) is expected to be large, while usually only a small number of labeled data for each fault class is available. From a statistical learning perspective, adopting a flat multi-class learning method and ignoring prior information will result in loss of valuable information, thus leading to degraded performance [42].

Secondly, the presence of different fault severity is well acknowledged in experiments but has long been ignored for FDD purpose. In a real building cooling system, faults naturally exhibit at various levels of severity due to different system/component degradations [43–46]. For instance, in the research of typical chiller faults, condenser fouling is a physical obstruction which is caused by the aggregation of non-decomposable chemical substances in the condenser tubes. It lowers the effective heat transfer coefficient and decreases the water flow rate in a manner consistent with the degree of aggregation. Hence the severity/degree of fault provides researchers/system managers valuable information to optimize maintenance actions, as well as to set priorities for different system scenarios. On the other hand, the advancement of the sensor network technology has greatly improved the capability to monitor temperature, flow rate, pressure, etc. with a refined spatial temporal granularity [44]. In short, detecting severity level in a data-driven framework is not only favorable, but also doable. Until now no work has tried to identify how serious the identified fault is.

In this paper we emphasize the importance of incorporating the prior knowledge of fault dependence (derived from chiller system characteristics). The objective of this paper is to design a novel data-driven FDD method, so as to: (1) recognize faulty working conditions and identify the fault type, (2) determine how serious the identified fault is. We propose a unified framework, namely Tree-structured Fault Dependence Kernel (TFDK) method to include the inter-class information and describe the fault severity levels. The tree-structured labels regard the severity levels as child nodes of each fault type rather than viewing them as independent classes. In addition, an on-line learning method is developed to train a multi-class classifier with streaming sensor measurement data. The effectiveness of TFDK has been evaluated on the experimental data of ASHREA Research Project RP-1043, and the results show significant improvement over the state-of-the-art approaches.

Compared with previous building FDD works, this paper presents its contributions in several ways. (1) A TFDK method is derived to make use of the fault dependence information, and thus achieving higher FDD accuracy compared with other methods; (2) an on-line learning method is developed to accommodate streaming data, which enables seamless integration to the BMS and sequential decision-making for HVAC schedule; (3) detailed information about the building performance is provided by identifying fault severity levels, hence providing researchers and building managers more options on taking actions to handle the faults.

In Section 2, we present the formulation of structured dependence information in the building cooling system. The derivation of TFDK which is based on structured building FDD formulation is given in Section 3. Section 4 presents the FDD results by TFDK and compares it with other multi-class classification methods. Section 5 summarizes the paper and suggests possible future work.