Applied Soft Computing 34 (2015) 402–425

Contents lists available at ScienceDirect

Applied Soft Computing

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

Unsupervised feature selection using swarm intelligence and consensus clustering for automatic fault detection and diagnosis in Heating Ventilation and Air Conditioning systems

Mitchell Yuwono [a,*], Ying Guo [b], Josh Wall [c], Jiaming Li [b], Sam West [c], Glenn Platt [c], Steven W. Su [a]

[a] Faculty of Engineering and Information Technology, University of Technology, Sydney (UTS), 15 Broadway, Ultimo, NSW 2007, Australia
[b] The Commonwealth Scientific and Industrial Research Organisation (CSIRO), Digital Productivity Flagship, Marsfield, NSW 2122, Australia
[c] The Commonwealth Scientific and Industrial Research Organisation (CSIRO), Energy Flagship, Mayfield West, NSW 2304, Australia

ARTICLE INFO

Article history:
Received 4 May 2014
Received in revised form 12 February 2015
Accepted 17 May 2015
Available online 22 May 2015

Keywords:
Consensus clustering
Feature selection
Ensemble Rapid Centroid Estimation (ERCE)
Particle Swarm Optimization
Fault detection and diagnosis
Heating Ventilation and Air Conditioning (HVAC) system

ABSTRACT

Various sensory and control signals in a Heating Ventilation and Air Conditioning (HVAC) system are closely interrelated which give rise to severe redundancies between original signals. These redundancies may cripple the generalization capability of an automatic fault detection and diagnosis (AFDD) algorithm. This paper proposes an unsupervised feature selection approach and its application to AFDD in a HVAC system. Using Ensemble Rapid Centroid Estimation (ERCE), the important features are automatically selected from original measurements based on the relative entropy between the low- and high-frequency features. The materials used is the experimental HVAC fault data from the ASHRAE-1312-RP datasets containing a total of 49 days of various types of faults and corresponding severity. The features selected using ERCE (Median normalized mutual information (NMI) = 0.019) achieved the least redundancies compared to those selected using manual selection (Median NMI = 0.0199) Complete Linkage (Median NMI = 0.1305), Evidence Accumulation K-means (Median NMI = 0.04) and Weighted Evidence Accumulation K-means (Median NMI = 0.048). The effectiveness of the feature selection method is further investigated using two well-established time-sequence classification algorithms: (a) Nonlinear Auto-Regressive Neural Network with eXogenous inputs and distributed time delays (NARX-TDNN); and (b) Hidden Markov Models (HMM); where weighted average sensitivity and specificity of: (a) higher than 99% and 96% for NARX-TDNN; and (b) higher than 98% and 86% for HMM is observed. The proposed feature selection algorithm could potentially be applied to other model-based systems to improve the fault detection performance.

1. Introduction

Heating Ventilation and Air Conditioning (HVAC) systems are important for maintaining the thermal comfort and indoor air quality at places such as offices, shopping malls, warehouses, schools, and homes [1,2]. According to the report by CSIRO [3], 25% of energy consumption in Australia is accounted from commercial buildings [3]. Moreover, HVAC systems represents 40–50% of energy use in these buildings [4]. In the United States (US), HVAC systems account for almost 31% of the electricity consumed by households [1]. Operational problems in the HVAC systems can cause excess energy consumption. Regular checks and maintenance are therefore crucial to prevent unnecessary consumption. However, due to the high reactionary maintenance costs, preventive or predictive maintenance practices are usually preferred to reactionary maintenance.

Discriminating a normally behaving HVAC system to a fault condition is a relatively well researched area. A variety of automatic fault detection and diagnosis (AFDD) techniques provide a number of benefits to the HVAC systems [5–7]. The current AFDD techniques available in the market for HVAC systems are mainly rule-based approaches [8–10], which obtain prior knowledge to derive a set of if-then-else rules and an inference mechanism that searches through the rule-space to draw conclusions. The rule-based systems can be based solely on expert knowledge (inferred from experience) or can be based on prior knowledge of a specific

* Corresponding author. Tel.: +61 430731938.
E-mail addresses: mitchellyuwono@gmail.com (M. Yuwono), Ying.Guo@csiro.au (Y. Guo), Josh.Wall@csiro.au (J. Wall), Jiaming.Li@csiro.au (J. Li), Sam.West@csiro.au (S. West), Glenn.Platt@csiro.au (G. Platt), Steven.Su@uts.edu.au (S.W. Su).

http://dx.doi.org/10.1016/j.asoc.2015.05.030
1568-4946/

system. Being one of the very first methods used in HVAC fault detection problems, the rule-based approaches have been most popularly used over the last decades.

Indeed the rule-based approaches come with advantages including ease of development, transparent reasoning, ability to reason even under uncertainty, and the ability to provide explanations for the conclusions reached. However, one must realize that most HVAC systems are installed in different buildings/environments. This generally means that rules or analytical models developed for a particular system cannot be easily applied to an alternative system. As such, the difficult process of determining and setting rules or generating analytical mathematical models must be tailored to each individual building/environment. The threshold method utilized in the rule-based system is prone to producing false alarms. Moreover, building conditions such as structure of the internal architecture design and even external factors (such as shading and the growth of plant life) often change after the system installation/initialization of a fault detection system, which can require rules/models that were originally appropriate to be revisited and updated. It can be learned that a number of weaknesses associated with this type of approach include the requirement of specific tailoring to a system, potential failure of the AFDD system due to its limited knowledge boundaries, and difficulty in updating the model when the AFDD system is installed in a different HVAC system. The aforementioned complications with the rule-based approach give rise to the data driven methods for AFDD in HVAC systems.

Regardless of the approach, the performance of an AFDD algorithm generally depends on the quality of the features. In CSIRO, we are developing a novel data-driven machine learning technique for AFDD in HVAC systems [4,11–14]. Preliminary results were presented in [11–14], showing the superior performance of the machine learning-based technique in detecting air-handling unit (AHU) faults to rule-based methods based on fault data obtained from ASHRAE Project 1312-RP up to 90% accuracy [13]. However, one limitation of the AFDD systems described in [11–13] is that they rely on features provided by field experts. As with rules, features that are particularly effective for a particular system may not guarantee equivalent performance when utilized in an alternative system.

Selecting the appropriate features is essential in any model-based frameworks. Feature selection aims for minimizing redundancies/mutual information between features such that the more important 'characteristic' features are not undermined. Specific faults exhibit specific symptoms which are observable only in certain clusters of features that behave differently to the others. The difficulty is that these cluster of features need to be constantly monitored as they may change dynamically depending on the condition of the HVAC system under investigation. Moreover, incorrect selections of these characteristic features are dangerous as they may adversely effect the final classifier to an extent that some obvious faults are overlooked. The motivation of this paper is therefore to design a reliable method for feature selection that can be used to augment the effectiveness of AFDD frameworks in general. The unsupervised data-driven feature selection algorithm is designed for HVAC systems operating under varying seasonal dynamics.

Evolutionary algorithms are particularly powerful for solving complex optimization problems with multiple local minima. For example, Differential Evolution (DE) has been used for optimization of pressure vessel structure design [15] and joint replenishment and distribution model [16]. Although the methods outlined in [15,16] are powerful for general purpose optimization, a major algorithmic restructuring is required to implement these algorithms for cluster optimization. Instead, our paper is interested in exploiting a lightweight evolutionary algorithm designed specifically for clustering purposes, the Rapid Centroid Estimation (RCE) [17].

Unsupervised feature selection based on data clustering is inherently an ill-posed problem where the goal is to group redundant features into some unknown number of clusters based on intrinsic information alone. For this paper, we utilize the Ensemble Rapid Centroid Estimation (ERCE) [17,18], a semi-stochastic multi-swarm clustering algorithm inspired by the Particle Swarm Optimization (PSO [19]), to determine the characteristic features for the specific season. The method is designed to automate the selection of characteristic features in each season. The block diagram of the proposed method is shown in Fig. 1.

The performance of the proposed feature selection algorithm was tested using two well established time-sequence classifiers: (a) Nonlinear Auto-Regressive Time Delay Neural Networks with Exogenous inputs (NARX TDNN); and (b) Hidden Markov Models (HMM) [13]. A comprehensive comparison would also be given with regards to other feature selection methods including Li's Manual selection [20], Complete Linkage (CL), Ensemble Evidence Accumulation K-means (EAC K-means) and Weighted Evidence Accumulation K-means (WEAC K-means).

The paper is structured as follows: Section 2 presents the overview of the proposed method as well as the materials used to examine its performance. Section 3 presents the detailed description for each component including feature extraction, feature selection, and the classifier used in experiment. Section 4 describes the theoretical foundations of the consensus clustering algorithm that we utilize for performing the feature selection. Section 5 describes the data utilized in the experiments. Section 6 presents a comprehensive experimental result of the proposed method and comparative analysis with other conventional feature selection and classification algorithms. Section 7 presents in depth analyses and discussion regarding the results. Finally, Section 8 presents the conclusion and future direction of the research.

2. General overview on HVAC systems

HVAC systems are configured and used to control the environment of a building or a zone including one or several rooms. The environmental variables may, for example, include temperature, air-flow, and humidity. The desired values/set-points of the environmental variables will depend on the intended use of the HVAC system. If the HVAC system is being used in an office building, the environmental variables will be set to make the building/rooms therein comfortable to humans. An HVAC system typically services a number of zones within a building. The system normally includes a central plant which includes:

- a hydronic heater and chiller,
- a pump system, which may include dedicated heated and chilled water pumps, circulates heated and chilled water from the heater and chiller through a circuit of interconnected pipes, and
- a valve system, which may include dedicated heated and chilled water valves, controls the flow of water into a heat exchange system (which may include dedicated heated and chilled water coils).

The heated and/or chilled water circulates through the heat exchange system before being returned to the central plant where the process repeats (i.e. the water is heated or chilled and recirculated). In the heat exchange system, energy from the heated/chilled water is exchanged with air being circulated through an air distribution system.

The HVAC system also includes a sensing system which typically includes a number of sensors located throughout the system, such