Energy and Buildings 116 (2016) 104–113

Contents lists available at ScienceDirect

Energy and Buildings

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

An improved fault detection method for incipient centrifugal chiller faults using the PCA-R-SVDD algorithm

Guannan Li[a], Yunpeng Hu[a,b], Huanxin Chen[a,*], Limei Shen[a], Haorong Li[c], Min Hu[a], Jiangyan Liu[a], Kaizheng Sun[a]

[a] Department of Refrigeration and Cryogenics Engineering, School of Energy and Power Engineering, Huazhong University of Science and Technology, 430074 Wuhan, China
[b] State Key Laboratory of Compressor Technology, Hefei General Machinery Research Institute, 230031 Hefei, China
[c] Durham School of Architectural Engineering and Construction College of Engineering, University of Nebraska-Lincoln, Omaha, NE 68182, USA

ARTICLE INFO

Article history:
Received 29 September 2015
Received in revised form
21 November 2015
Accepted 24 December 2015
Available online 29 December 2015

Keywords:
Chiller
Fault detection
Principal component analysis
Residual
Severity level
Support vector data description

ABSTRACT

Detecting the faults at the incipient stage is important for keeping chiller systems healthy and saving energy and maintenance cost. Traditional principle component analysis (PCA) and support vector data description (SVDD) methods are insensitive to two common faults, condenser fouling (CdF) and refrigerant leakage (RfL). To improve the fault detection performance, this study proposed a PCA-R-SVDD based method. Instead of principle component subspace (PCs), it develops a SVDD model in the residual subspace (Rs) using the PCA modeling residual data. The SVDD based distance based monitoring statistic was used for fault detection. The proposed method shows significant improvement comparing with the traditional methods due to the better fault data distribution and tighter monitoring statistic. It is sensitive to six common faults. At least 50% of the fault data can be correctly detected even at the least severe fault level. Centrifugal chiller experimental data from the ASHRAE Research Project 1043 (RP-1043) was used to evaluate the methods.

1. Introduction

Chillers are commonly used for providing sufficient chilled water to remove the additional heat in the air conditioned buildings. They are large energy consumers which account for more than 40% of the total energy used in the commercial and industrial buildings for space conditioning [1,2]. Due to the wide operating range and good part load characteristic, the chiller system can meet the cooling load demand even at fault conditions. But faults can lead to significant energy efficiency reduction to the system, sometimes the efficiency drop can be as much as 30% [2]. Survey results indicate that condenser fouling (CdF) and refrigerant leakage (RfL) are two of the most frequently occurring and costly faults for centrifugal chillers [3,4]. Even a slight CdF fault may cause great efficiency loss to the system at some operating conditions. It is also required that the detection of refrigerant leakage (RfL) fault should be as soon as possible in order to avoid refrigerant emission and further environmental harm. Besides, the early detection of these faults can reduce the occurring frequency of costly compressor failures. Therefore, it is necessary for a fault detection method to be sensitive to the faults occurring at the incipient stage.

Over the past two decades, many researchers have devoted themselves on applying various methods for detecting the chiller faults [4–8]. As reviewed in [5], there are two main categories of fault detection methods: one is the physical model based methods [4,6,9] and the other is the data based methods [7,8]. Physical models contribute a lot to explaining the fault mechanism. But they are very complex and time consuming. In addition, these models may be difficult to give precise description of the real chiller system working at a wide range of operational conditions, especially at the off-design conditions. On the contrary, the data based methods can overcome these limitations by developing models using operating data of the real chiller system covering a wider range of working conditions. Consequently, studies on the data based approaches for chiller fault detection have caught much research attention in recent years.

Principal component analysis (PCA) is one of the most popular data based methods. It has been successfully applied for sensor fault detection [10–12] because of its superior performance on extracting the system changing information [10]. Yet this method is not sensitive to some component or system faults. Zhao [8] performed the PCA method for Cdf and RfL fault detection using the

* Corresponding author.
E-mail address: chenhuanxin@tsinghua.org.cn (H. Chen).

http://dx.doi.org/10.1016/j.enbuild.2015.12.045

Nomenclature

a	SVDD hypersphere center
C	penalty weight
CPV	cumulative percent variance
D	SVDD distance based distance statistic refers the distance to hypersphere center
$D(z_i)$	distance of a test data point z_i to the center of the SVDD hypersphere
E	residual matrix
f	objective function
g	Gaussian kernel width parameter in the SVDD model
$K(x_i, x_j)$	Kernel function
m	dimension of training data matrix
min	minimum function optimization problem
n	number of data points
P	loading matrix in the principal component subspace
$\tilde{P}$	loading matrix in the residual subspace
Q	Q statistic value
R	SVDD hypersphere radius and the threshold for the distance based statistic
S	feature space
$s.t$	subject to constrain conditions
T	score matrix in the principal component subspace
t	score vector in the principal component subspace
$\tilde{T}$	score matrix in the residual subspace
T^2	T^2 statistic value
U	temperature measurement
X	training data matrix
x_i	training data point
Z	testing data set
z_i	test data point
$\hat{z}$	modeled part of z
$\tilde{z}$	un-modeled part of z
α_i	Lagrange multipliers
γ_i	Lagrange multipliers
λ_i	eigenvalues
η_i	cross-validation accuracy for the ith subset
ξ_i	slack variable in the SVDD model
ϕ	nonlinear map
Λ	eigenvectors matrix

Subscripts

PC	principal component
PCs	principal component subspace
Rs	residual subspace
α	threshold with a given level of significance

experimental data from the ASHEAR Research Project 1043 (RP-1043) [4]. No more than 3% of the fault data were correctly detected for the two faults at their least severe levels. The fault detection results may be even worse if the noisy practical operating data are used [13].

Support vector data description (SVDD) is a powerful one-class classification algorithm. This pattern recognition based classification approach performs well in describing the non-linear industrial process data violating the Gaussian distribution [14,15]. Many researchers have tried the SVDD algorithm to obtain improved fault detection performance. Using the kernel method, SVDD develops a tight-boundary hypersphere to separate the normal data from the fault data. But the SVDD based fault detection ratios for CdF and RfL faults are merely 28% and 34% at the first severity level, respectively. The ratios are still relatively low for the two frequently occurring and costly chiller faults. More energy may have been wasted before the faults gradually develop to the more severe level that can be precisely detected.

However, the two algorithms have complement advantages. Many researchers [16,17] have tried to combine the two methods. Jiang [16] improved the sensor fault detection result in chemical process by the PCA-PC-SVDD[1] method, which integrates the SVDD method with the PCA method. Instead of using the traditional PCA method, Liu [17] developed a SVDD model in the principal component (PC) subspace to achieve better fault detection performance for industrial process. The PC data was used to capture the information illustrating the condition changes to the chillers. But chiller systems usually work at a wide range of operating conditions, this method may yield bad results since it may be not able to correctly distinguish between the normal data of other operating conditions and the fault data [18]. Another fault detection method is to check whether the modeling residuals are out of the confidence intervals or not [19,20]. But the traditional monitoring statistic PCA used in the residual subspace (Rs) is insensitive to chiller component faults [8]. So the current study proposed a new PCA-R-SVDD[2] method, which established the SVDD based distance statistic in the residual subspace. Four data based methods, PCA, SVDD, PCA-PC-SVDD and PCA-R-SVDD, were applied for chiller fault detection. The chiller experimental data from ASHEAR Research Project 1043 (RP-1043) was used to analyze the fault detection performance. The fault detection sensitivity to faults at incipient stage were also evaluated.

2. Outline of the PCA-R-SVDD algorithm

2.1. Principal component analysis (PCA)

PCA is an effective multivariate statistical algorithm and has been widely used for dimension reduction and feature extraction [21]. PCA transforms a group of correlated variables into a new group of variables which are uncorrelated or orthogonal to each other. Given a training data matrix X, it can be separated into two subspaces according the PCA method as shown in Eq. (1),

$$X = TP^T + E = TP^T + \tilde{T}\tilde{P}^T \tag{1}$$

where, T and P are score and loading matrices in the PC subspace, respectively. E is the residual matrix and can be further decomposed into two parts in the Rs, the residual score matrix $\tilde{T}$ and the residual loading matrix $\tilde{P}$. The score matrix T and the residual score matrix $\tilde{T}$ are the representations of X in the PC subspace and the Rs, respectively. The loading matrix P determines the projection relationship between the training data matrix X and corresponding score matrix T. While the residual loading matrix $\tilde{P}$ determines that between the training data matrix X and the corresponding residual score matrix $\tilde{T}$.

In Fig. 1, the new input vector z is mapped into the PC and residual subspaces respectively.

$$z = \hat{z} + \tilde{z} \tag{2}$$

where, $\hat{z}$ is the modeled part in PC subspace. $\tilde{z}$ is the un-modeled part in Rs. The PC subspace has the potential to capture the process systematic variations. Whereas the Rs represents the random noise or even some abnormal information [10]. As shown in Eqs. (3) and

[1] In this paper, we called it the PCA-PC-SVDD method since the SVDD model was constructed in the PC subspace based on the PCA decomposition process.

[2] In this paper, we called it the PCA-R-SVDD method since the SVDD model was constructed in the residual subspace based on the PCA decomposition process.