

Fault identification for chiller sensor based on partial least square method

Bang Wu¹, Yunpeng Hu^{2,*}, Chuanhui Zhou¹, Guaiguai Chen¹, and Guannan Li¹

¹ School of Urban Construction, Wuhan University of Science and Technology, Wuhan, Hubei, PR China

² Department of Building Environment and Energy Engineering, Wuhan Business University, Wuhan, Hubei, PR China

Abstract. Sensor failures can lead to an imbalance in heating, ventilation and air conditioning (HVAC) control systems and increase energy consumption. The partial least squares algorithm is a multivariate statistical method, compared with the principal component analysis, its compression factor score contains more original data characteristic information, therefore, partial least squares have greater potential for fault diagnosis than the principal component analysis. However, there are few studies based on partial least squares in the field of HVAC. In order to introduce partial least squares into the field, based on the partial least squares fault detection theory, a fault analysis method suitable for this field is proposed, and the RP1403 data published by ASHARE was used to verify this method. The results show that on the basis of selecting the appropriate number of principal components, partial least squares have the ability to diagnose the fault of the chiller sensor. With the known fault source, partial least squares regression, a method with better data reconstruction accuracy than principal component analysis, is used to repair the fault. Finally, the purpose of fault identification can be achieved.

1 Introduction

The sensor of the automatic control system is an important component to ensure the stable operation of the HVAC. The failure of the sensor will cause the HVAC system to deviate from the normal working state and lead to more energy consumption. The chiller is the main energy supply equipment of the HVAC system, so it is of great significance to carry out the research on the sensor failure of the chiller.

Most recent research on fault detection and diagnosis (FDD) of chiller sensor is based on data analysis methods. The principal component analysis (PCA) method can describe data well through dimensionality reduction, and the algorithm is fast, so that it is the most widely used in the HVAC field for fault diagnosis. Zhang, N et al [1] used the improved KPCA for fault detection, and proved that the accuracy of this method is superior to PCA. Hu Y et al [2] proposed a self-adaptive PCA method, which greatly improved the efficiency of PCA-based chiller sensor failure detection.

PLS and PCA are two basic multivariate statistical algorithms. In fact, PLS is also an improved algorithm based on PCA. They have a lot in common, such as the ability to reduce dimensionality and describe data well. PLS model connects input and output through several dimensionless score vectors, and builds the model by mining the maximum correlation between input and output, which can be used to monitor input variables or to

predict output variables. Feng L V et al [3] applied PLS to the fault diagnosis of wind turbine, indicating that compared with PCA, PLS makes full use of sample space information and can perform fault diagnosis more effectively. A. Chen et al [4] used PCA and PLS in the fault process of the wastewater treatment, and concluded that both could achieve the purpose of the fault diagnosis. S W Choi et al [5] conducted fault diagnosis on the chemical process based on multi-block partial least square method, which could find fault blocks or fault variables.

It can be seen that the PLS algorithm is also widely used in the field of fault diagnosis, and has its own advantages compared with PCA. In fact, the special function of PLS supporting multiple output variables [6] makes PLS have greater potential for fault source diagnosis. However, in the field of HVAC, there are fewer studies based on PLS. In this paper, a PLS-based sensor fault identification method for chiller is proposed into the field of HVAC.

2 Materials and Methods

In order to find a PLS model specifically for the fault analysis of chiller sensor, it is necessary to analyze the fault monitoring theory of PLS in detail. Then, according to the characteristics of the chiller, the fault diagnosis and data reconstruction strategy are formulated. The experimental data of the RP1043 is selected to verify the

* Corresponding author: YunpengHu@wbu.edu.cn

proposed method.

2.1 Overview of PLS fault detection

PLS combined with KPI refers to the establishment of the input and output relationship between the process variables and the key performance indicator (KPI) of the industrial process. Use Q , T^2 statistics to detect failures of input variables, and feedback whether these failures are related to KPI, at the same time, it can realize the online prediction of KPI [7].

However, the score vectors of PLS is not completely output related. As a result, the utilized test statistics yet fault diagnostic results offered by the standard approach are problematic for KPI-related process monitoring. When applied to HVAC equipment, all the selected variables are related to each other due to the high coupling of the HVAC system, then the Q and T^2 statistics are only used for fault detection, not for judging whether the fault is KPI related. The fault detection idea when PLS is applied in HVAC system is shown in Figure 1, this is a single output PLS model. We keep the prediction of KPI, the failure of KPI variable can be detected by comparing predicted and measured values. The Q statistic and Q contribution rate can be used for fault identification of input variables, this is what this paper will show in detail.

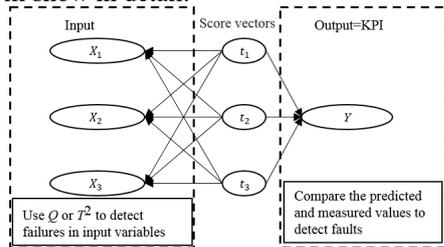


Fig. 1. Schematic diagram of HVAC system fault detection based on PLS

2.2 SIMPLS algorithm

Simple partial least square (SIMPLS) method [8] has the advantages of faster calculation and easy explanation, compared with the traditional nonlinear iterative partial least square (NIPALS) method. Using NIPALS or SIMPLS does not affect the results of fault identification. Because of the above advantages, SIMPLS was adopted. Given a training data $X \in \mathfrak{R}^{n \times m}$, $Y \in \mathfrak{R}^{n \times z}$, centralized to get $X_0 \in \mathfrak{R}^{n \times m}$, $Y_0 \in \mathfrak{R}^{n \times z}$, the matrix solution can be obtained under the following three conditions.

$$Cov(t_1, u_1) = \sqrt{Var(t_1)Var(u_1)}Corr(t_1, u_1) \rightarrow \max \quad (1)$$

$$t_\beta = X_0 w_\beta, u_\beta = Y_0 c_\beta \quad (2)$$

$$\|w_\beta\| \rightarrow 1, \|c_\beta\| \rightarrow 1, t_\beta^T t_\beta = 0, \psi > \xi \quad (3)$$

The implementation of SIMPLS algorithm needs to go through the following steps:

- 1) Write $S=X_0^T Y_0$
- 2) Perform SVD decomposition on S , where w is the first column of the left singular matrix of S

- 3) Calculate the scores: $t=X_0 w$, calculate the norm of t : $normt=\sqrt{t^T t}$
- 4) Normalize t : $t_{norm}=t/normt$
- 5) Adapt weights accordingly: $w_{adapt}=w/normt$
- 6) Calculate the load of X block: $p=X_0^T t_{norm}$
- 7) Store w_{adapt}, t_{norm}, p into W, T, P respectively
- 8) Deflate S : $S = P^\perp S = S - P(P^T P)^{-1} P^T S$
- 9) Cycle 2)- 8) find all SIMPLS factors

The number of score vectors of the PLS algorithm will affect the fault diagnosis performance, this study refers to the method described in Reference[6] to determine, suppose it is L_s . According to the idea of principal component regression, the following basic model of PLS is obtained:

$$X_0 = TP^T + E_{Final}, Y_0 = TD^T + F_{Final} \quad (4)$$

The regression coefficient B_{pls} is calculated as follows:

$$B_{pls} = WD^T = W(T^T Y_0) \quad (5)$$

It can be seen that the SIMPLS algorithm extracts the characteristic information of the data by adjusting S , and each time the first left singular vector is taken from the remaining information as the projection axis, the PCA selects the eigenvectors corresponding to multiple eigenvalues from large to small. Compared with PCA, PLS makes full use of the information contained in the data. Therefore, it has greater potential for fault diagnosis and data reconstruction than PCA. In addition, B_{pls} can be used for prediction of output variables, this function is called PLSR.

2.3 Sensor fault identification based on Q statistic of PLS

The Q statistic of PLS realizes the detection of input variables failure. When the single-output PLS model was chosen, the detection range of Q statistic was maximum and the model was simplified.

The PLS-based sensor fault identification process adopted is: fault detection, fault diagnosis, data reconstruction, and fault identification, hereinafter referred to as FDDRI.

2.3.1 Fault detection

Use Q statistic to detect failures. The $Q(i)$ statistic is the squared prediction error of the i -th sample, expressed as:

$$Q(i) = \sum_j^m (X_{0(i,j)} - X_{0(i,j)} W P^T)^2 \quad (6)$$

When the confidence is $\alpha=0.05$, the Q statistic threshold Qa is derived from the training data and can be calculated by the following formula:

$$Qa = g \chi_{h,\alpha}^2 \quad (7)$$

Where g is the weighting parameter, h is the degree of freedom. When the Q statistic is higher than the threshold Qa , there must be on the fault condition.

2.3.2 Fault source diagnosis

Q contribution rate graph is used for fault source diagnosis.

$$Q_{con}^j = e_j^2 / Q \quad (8)$$

Where: $\sum_{j=1}^m Q_{con}^j = 1$, when a certain data in the measurement data set has a fault, then the j -th component of the sampled data has changed, resulting in a deviation of the residual vector e of the data in the j -th dimension. This deviation affects the calculation of the Q statistic, it also leads to an increase in the Q contribution rate of the dimension. Therefore, by determining the dimension j where the maximum contribution rate is located, the j -th sensor can be determined as the source of the fault.

2.3.3 Data reconstruction and fault identification

After finding the fault source using the Q contribution rate graph, PLSR is used to repair faulty data. Assuming that the j -th sensor fails, the fault data reconstruction algorithm is expressed as:

$$X_{j,rc} = PLSR(X_{except(j)}) \quad (9)$$

In the formula, $X_{j,rc}$ represents the reconstructed value of the j -th input variable sensor. $X_{except(j)}$ means sensor data of other input variables except the j -th sensor. The meaning of (9) is to use healthy input variable sensors data to reconstruct the faulty input variable sensor data. The PLSR symbol in (9) indicates that this process is implemented using partial least squares regression.

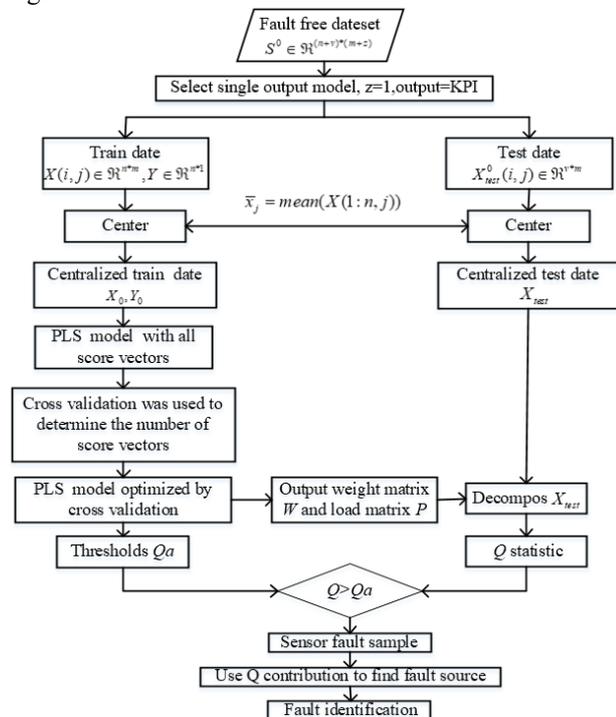


Fig. 2. FDDRI flowchart based on the Q statistic of the PLS

After data reconstruction, the reconstructed data matrix is obtained, and the Q statistics of the reconstructed data matrix are calculated. If the Q statistic of the reconstructed data is in a normal state, then the fault is correctly identified. Suppose the number of test

data samples is v . the flowchart of fault identification based on the Q statistic of the PLS algorithm is shown in Figure 2. The entire process is based on Q statistic, and then the FDDRI range includes all input variables.

2.4 Feature variable selection and model establishment

According to the heat balance and the laws of thermodynamics, nine characteristic variables are selected from the RP1043 data set. The nine variables correspond to nine important sensors, and these variables are shown in international units in Table 1.

Table 1. Characteristics of the variable

Characteristic variable name	Marks
Temperature of leaving evaporator water(°C)	TEO
Temperature of entering evaporator water(°C)	TEI
Evaporator water flow rate (m ³ /h)	FWE
Temperature of leaving evaporator water (°C)	TCO
Temperature of entering evaporator water (°C)	TCI
Condenser water flow rate (m ³ /h)	FWC
Condenser pressure (Gage pressure, KPa)	PRC
Evaporator pressure (Gage pressure, KPa)	PRE
The chiller power (KW)	W _e

Among the various parameters of the chiller, W_e is regarded as the key performance indicator (KPI) of the system. According to the fault detection ideas mentioned above, the fault detection model data matrix of PLS is:

$$X=[TEO,TEI,FWE,TCO,TCI,FWC,PRC,PRE], Y=[W_e].$$

The FDDRI analysis matrix based on its Q statistic is :[TEO,TEI,FWE,TCO,TCI,FWC,PRC,PRE].

In the FDDRI process, 828 sets of representative data are selected. The first 415 sets of date are healthy training data, the last 413 sets of date are used as test data, and the fault is artificially added in the test data.

3 Results & Discussion

3.1 Results of FDDRI

415 samples of training data are used for model training and model optimization. The number of score vector is determined to be four, the Q statistic threshold $Q_a=3.976$ is obtained, and the detailed results of fault identification are shown below. After introducing a +2°C deviation fault to the TEO sensor, the Q statistics of the training data and the test data are shown in Figure 3. 99.52% of the Q statistics of all test data are higher than the threshold, the fault is detected, and this indicator is only 6.27% in the training data, the result is credible.

The Q contribution rate of this deviation fault is shown in Figure 4. In the Q contribution rate graph, the first 415 sets of data are training data, and the contribution of FWE is the largest. When the fault is introduced (after 416 sets), the contribution rate of TEO changes most obviously, and the Q contribution rate of TEO is dominant. It is considered that the fault occurred

on the TEO sensor, but the degree of the fault is unknown. According to the aforementioned data reconstruction method, after the TEO sensor is diagnosed as the source of the fault, reconstructed data and its Q statistic are calculated. The Q statistic and the Q statistic of the reconstructed data under a TEO sensor deviation fault condition of $+2^{\circ}\text{C}$ are shown in Figure 5. In the reconstructed test data, the Q statistics exceeding Q_a accounted for 8.96% of the total test data. However, this proportion is also 6.27% in the training data, indicating that the reconstructed data is in a normal condition, which verifies the effectiveness of the data reconstruction method.

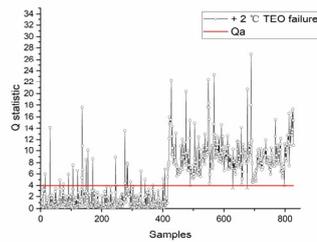


Fig. 3. The Q statistic graph of TEO $+2^{\circ}\text{C}$

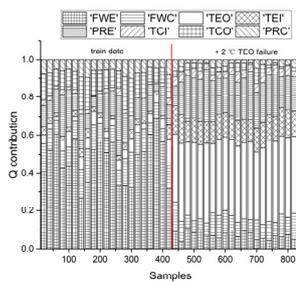


Fig. 4. Q contribution rate of TEO $+2^{\circ}\text{C}$

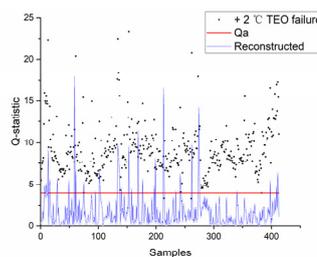


Fig. 5. TEO $+2^{\circ}\text{C}$ Q statistic before and after reconstruction

For different sensors, the Q statistic of PLS has different fault detection sensitivity, Q contribution rate chart can correctly diagnose six types of sensor failures. After the fault source is correctly diagnosed, the Q statistics of the reconstructed data all become normal.

3.2 Discussion

PLS and PCA are two basic algorithms of multivariate statistics. They have many points in common, and the PCA method is the most widely used in the fault diagnosis of chiller sensors. Therefore, it is of great significance to compare PLS and PCA.

PLS and PCA have many common advantages, for

example, they can describe data well and are easy to use. From the perspective of algorithm principle, PLSR is more suitable for data reconstruction, so the PLS-based method is better in data reconstruction. The compression factor score of PLS contains more characteristic information than that of PCA, so it has greater potential in failure analysis and has sufficient improvement methods. However, the fault diagnosis based on Q statistic of PLS does not include output variable. Therefore, other methods are needed to diagnose the fault of the output variable.

4 Conclusions

The research results show that the PLS-based chiller sensor fault identification method proposed in this paper is effective. The introduction of PLS into the HVAC field is significant. Similar to PCA, it can describe data characteristics well and is easy to use. In addition, PLS have capabilities that PCA does not. PLS supports mutual reconstruction from multi-dimensional matrix to multi-dimensional matrix, and will have the potential for fault diagnosis of multiple fault sources, which can supplement the defect of Q statistics. Optimizing data quality by wavelet decomposition or using other improved methods based on PLS can greatly improve the efficiency of fault detection. The FDD potential of PLS algorithm needs to be further explored and studied in detail.

Acknowledgments

This work is supported by Excellent Young and Middle aged Scientific and Technological Innovative Team in Universities of Hubei, China (T201829).

References

1. Zhang, N., Gao, X., Li, Y., Wang, P.(2016) Fault detection of chiller based on improved KPCA. IEEE, 2951-2955.
2. Hu, Y., Chen, H., Xie, J., Yang, X., Zhou, C. (2012) Chiller sensor fault detection using a self-Adaptive Principal Component Analysis method. Energy and buildings, 54:252-258.
3. Feng, L.V., Zhang, Z., Zhai, K., Ju, X.(2018) Research on fault diagnosis method of wind turbine based on partial least square method. IEEE, 925-928.
4. Chen, A., Zhou, H., An, Y., Sun, W. (2016) PCA and PLS monitoring approaches for fault detection of wastewater treatment process. IEEE, 1022-1027.
5. Choi, S.W., Lee, I. (2005) Multiblock PLS-based localized process diagnosis. Journal of process control, 15: 295-306.
6. Wold, S., Sjöström, M., Eriksson, L. (2001) PLS-regression: a basic tool of chemometrics. Chemometrics and intelligent laboratory systems. 58:109-130.

7. Yin, S., Ding, S.X., Xie, X., Luo, H. (1982) A Review on Basic Data-Driven Approaches for Industrial Process Monitoring. *IEEE transactions on industrial electronics*, 61:6418-6428.
8. De Jong, S. (1993) SIMPLS: An alternative approach to partial least squares regression. *Chemometrics and intelligent laboratory systems*, 18: 251-263.