

Research Article

Fault Diagnosis Method Based on Information Entropy and Relative Principal Component Analysis

Xiaoming Xu and Chenglin Wen

School of Automation, Hangzhou Dianzi University, Hangzhou 310018, China

Correspondence should be addressed to Chenglin Wen; wenc1@hdu.edu.cn

Received 19 December 2016; Accepted 22 January 2017; Published 20 February 2017

Academic Editor: Xiao He

Copyright © 2017 Xiaoming Xu and Chenglin Wen. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In traditional principle component analysis (PCA), because of the neglect of the dimensions influence between different variables in the system, the selected principal components (PCs) often fail to be representative. While the relative transformation PCA is able to solve the above problem, it is not easy to calculate the weight for each characteristic variable. In order to solve it, this paper proposes a kind of fault diagnosis method based on information entropy and Relative Principle Component Analysis. Firstly, the algorithm calculates the information entropy for each characteristic variable in the original dataset based on the information gain algorithm. Secondly, it standardizes every variable's dimension in the dataset. And, then, according to the information entropy, it allocates the weight for each standardized characteristic variable. Finally, it utilizes the relative-principal-components model established for fault diagnosis. Furthermore, the simulation experiments based on Tennessee Eastman process and Wine datasets demonstrate the feasibility and effectiveness of the new method.

1. Introduction

In the process of the industry manufacturing, there is a large amount of variables that is highly correlative; these variables contain the essential information that would be helpful to judge the status of the system. As a result, it is an important problem to find and predict the fault through this information to ensure that the equipment always works in a safe and reliable way [1, 2].

However, during the industry manufacturing, the collected characteristic variables have different units; this raises a problem that we may come out with the different result only due to the unit difference; hence, we have to standardize the unit. What is more, after the standardization, it is inevitable to lose the diversity among different variables and present the property of distribution uniformity in the perspective of geometry which makes it hard to extract the principle component for compression and diagnosis. As to overcome these problem, some methods have been proposed recently [3–8]. Shi et al. use the Mahalanobis distance for relative transformation to reduce the effect of the dimension standardization

[4]. Tang et al. propose a relative transformation principal component analysis to reduce the data noise for the transformation oil breakdown voltage prediction [5]. Yi et al. introduce a relative transformation operator to change the original variables in the spatial distribution and eigenvalues of the covariance matrix in the feature space [6]. Wen et al. propose a method called Relative Principle Component Analysis (RPCA); it introduces weighting for each variable based on the prior information of the system to eliminate the false information due to standardizing the variable units [7, 8], but the shortage of this method is that it needs a large amount of prior information from the system which is hard to gain in real engineering application.

In order to solve the problem, this paper introduces the concept of information entropy and proposes a new fault diagnosis method that combines the information entropy and relative transformation PCA which is called information entropy relative transformation PCA (InEnRPCA). The information entropy is put forward by Shannon in 1948 [9]; it indicates that the redundancy exists among any information and can be measured based on the symbol in the information such

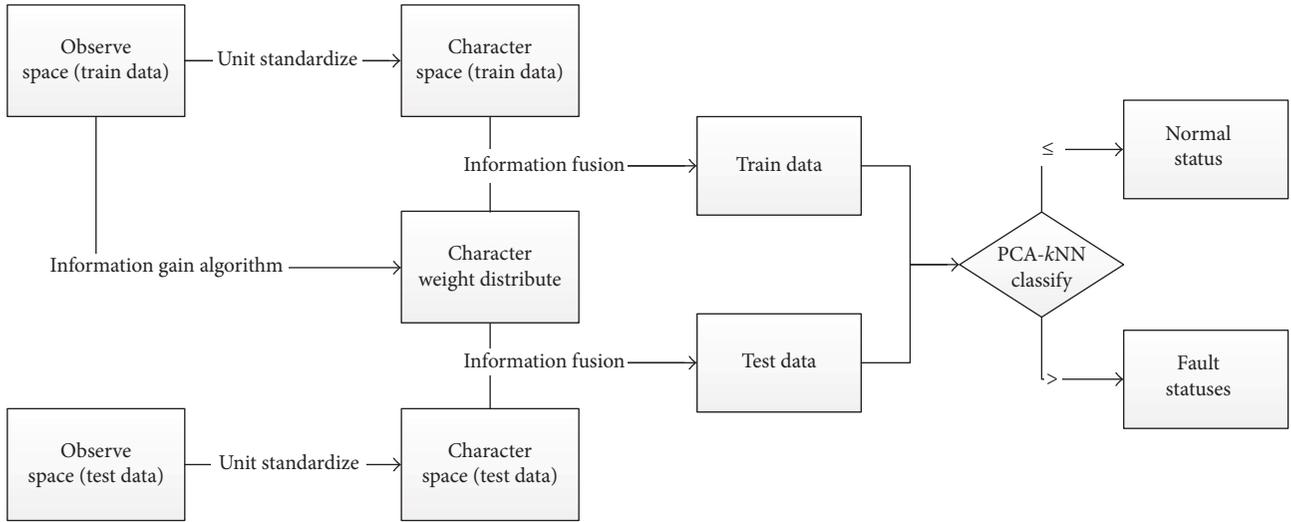


FIGURE 1: InEnRPCA algorithm process.

as the number, alphabet, and words. With the development of the information theory, the information entropy has become an effective method to measure the degree of importance for each feature in the sample and achieved a wide application in many areas. For instance, while using the decision-making tree for classification, Hu et al. use the information entropy to calculate the significance for each feature and then prune the decision-making tree and reduce the false alarm [10]. Y. Y. Chen and Y. M. Chen use the information entropy to measure the uncertainty on the data in the attribute reduction algorithm [11]. Wang et al. use the information entropy to balance the weight of each sememe in the area of natural language processing [12].

As to the problem we are facing, we start in the perspective of information theory and use the information gain algorithm to extract information entropy as the heuristic knowledge from the original dataset, then use it to calculate the relative transformation factor and allocate the weight for each standardized characteristic variable, and, finally, use the corresponding RPCA method for fault diagnosis.

The rest of this paper is organized as follows: in Section 2, we review the information entropy and information gain definition and algorithm. The original relative transformation PCA method is given in Section 3. Our simulation experiments on Tennessee Eastman process and the Wine dataset from UCI are stated in Section 4, where we make a comparisons between PCA, USPCA, and improved InEnRPCA with thirteen datasets to demonstrate the effectiveness of the new method. Finally, Section 5 gives conclusions and some discussions.

2. Overview of Our Approach

A brief overview of our fault diagnosis approach is given in this section. As the framework in Figure 1 shows, the proposed approach consists of two parts: calculating the relative

transformation operators based on information entropy and fault diagnosis based on RPCA- k NN.

In the part for calculating the relative transformation operators, information entropy and information gain algorithm are applied on train data to get the relative transformation operators. Then, the operators are combined with the original data to get the relative transformation matrix.

In fault diagnosis part, the RPCA are used to deal with processed data for dimension reduction, and k NN are used for classification training. After the above process, a model is built for further fault diagnosis.

3. Information Entropies

In the subject of the information theory, probability, and statistics, entropy is used to describe the uncertainty of random variable and can be used to show the reduced degree of the information uncertainty for set Y after getting the character X .

3.1. The Definition of the Information Entropy and Information Gain

Theorem 1 (information entropy). Assume M is Lebesgue measure set generated by measurable set S with algebra σ and measure μ where $\mu(M) = 1$, also M can be shown in the form of the incompatible sets with $A = A_i$, that is, $M = \bigcup_{i=1}^n A_i$ and $A_i \cap A_j = \Phi$, $\forall i \neq j$ [13]. Then, we can come to a conclusion:

$$H(A) = - \sum_{i=1}^n \mu(A_i) \log \mu(A_i), \quad (1)$$

where $\mu(A_i)$ is the measure for A_i .

Theorem 2 (information gain). Given the train set D and the corresponding feature A , the empirical entropy $H(D)$ means the

uncertainty level to classify set D , and the conditional empirical entropy $H(D | A)$ means the uncertainty level to classify feature A in the condition of set D . Then, the difference between them is the information gain which stands for the reduced level of the uncertainty to classify the set D by given feature A .

Obviously, as to the dataset D , the information gain depends on the features and different features have different information gain, the bigger the information gain, the stronger the ability it has for classifying.

3.2. The Information Gain Algorithm. Assume the train dataset is D ; $|D|$ denotes the sample capacity which is equal to the number of the samples. There are K classes in set D with each name being C_k ($k = 1, 2, \dots, K$) and $|C_k|$ is equal to the number of the samples which belongs to C_k ; that is, $\sum_{k=1}^K |C_k| = |D|$. Also, assume there are n values in feature A called $\{a_1, a_2, \dots, a_n\}$ and then divide set D into n parts with each name being D_i ($i = 1, 2, \dots, n$) and $|D_i|$ is equal to the number of the samples that belongs to D_i ; that is, $\sum_{i=1}^n |D_i| = |D|$. Based on the above, define the set D_{ik} as the intersection of class C_k and subset D_i , that is, $D_{ik} = D_i \cap C_k$, and $|D_{ik}|$ is the number of D_{ik} [14]. Then, the information gain algorithm can be shown as follows: input: train dataset D and feature A ; output: the information gain of set D by feature A defined as $g(D, A)$.

Step 1. Calculate the empirical entropy of dataset $H(D)$:

$$H(D) = - \sum_{k=1}^K \frac{|C_k|}{|D|} \log_2 \frac{|C_k|}{|D|}. \quad (2)$$

Step 2. Calculate the empirical entropy of the feature A in the condition of set D :

$$\begin{aligned} H(D | A) &= \sum_{i=1}^n \frac{|D_i|}{|D|} H(D_i) \\ &= - \sum_{i=1}^n \frac{|D_i|}{|D|} \sum_{k=1}^K \frac{|D_{ik}|}{|D_i|} \log_2 \frac{|D_{ik}|}{|D_i|}. \end{aligned} \quad (3)$$

Step 3. Calculate the information gain which is also the relative transformation operator M_A :

$$\begin{aligned} g(D, A) &= H(D) - H(D | A), \\ M_A &= g(D, A). \end{aligned} \quad (4)$$

Step 4. Repeat the above processes from Steps 1 to 3 for each feature i in the sample and get the relative transformation operator with each feature:

$$M_i = g(D, i). \quad (5)$$

4. Relative Transformation Principle Component Analysis

After the standardization for the data dimensions by the traditional PCA method, it may bring in some fake information to the principle element due to the uniform distribution.

Therefore, we use the relative transformation PCA method to solve the problem [15]. Below is the brief RPCA process.

Assume $X \in R^{a \times b}$, where a corresponds to the number of the samples and b is the number of features.

Step 1. Transfer the original data to the standardized data with mean zero and variance one.

The mean for each column is

$$m = \frac{1}{a} (x^0)^T l_a, \quad (6)$$

where $l_a = [1, 1, \dots, 1]^T \in R^a$; then,

$$X = (X^0 - l_a m^T) \Sigma^{-1}, \quad (7)$$

where Σ_i ($i = 1, \dots, b$) is the standard deviation for each column.

We set the relative transformation:

$$\begin{aligned} X^R &= XM \\ &= \begin{bmatrix} x_1(1) & x_1(2) & \cdots & x_1(N) \\ x_2(1) & x_2(2) & \cdots & x_2(N) \\ \vdots & \vdots & \ddots & \vdots \\ x_n(1) & x_n(2) & \cdots & x_n(N) \end{bmatrix} \\ &\quad \times \begin{bmatrix} M_1 & 0 & \cdots & 0 \\ 0 & M_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & M_n \end{bmatrix} \\ &= \begin{bmatrix} x_1^R(1) & x_1^R(2) & \cdots & x_1^R(N) \\ x_2^R(1) & x_2^R(2) & \cdots & x_2^R(N) \\ \vdots & \vdots & \ddots & \vdots \\ x_n^R(1) & x_n^R(2) & \cdots & x_n^R(N) \end{bmatrix}, \end{aligned} \quad (8)$$

where X^R is relative transformation matrix from X and M is a diagonal matrix in which each of the variables M_i is the relative transformation operator that can be achieved from Section 3.2.

Step 2. Compute the covariance matrix R_{X^R} from X^R :

$$R_{X^R} = \frac{X^{RT} X^R}{a - 1}. \quad (9)$$

Step 3. Compute the eigenvalue λ and the corresponding eigenvector of the covariance matrix R_{X^R} . Assume all eigenvalues satisfy $\lambda_1 \geq \lambda_2 \geq \dots$, where λ_i stands for the i 'th eigenvalue and p_i is the corresponding eigenvector to λ_i .

$$\begin{aligned} |\lambda I - R_{X^R}| &= 0, \\ |\lambda_i I - R_{X^R}| p_i &= 0; \quad i = 1, 2, \dots, b. \end{aligned} \quad (10)$$

TABLE 1: The datasets descriptions and parameter constitution.

Dataset	Numbers		Dimensions	Classes
	Train	Test		
TE (1–4)	800	400	52	4
TE (5–8)	800	400	52	4
TE (9–12)	800	400	52	4
TE (13–16)	800	400	52	4
TE (17–20)	800	400	52	4
TE (1–3)	600	300	52	3
TE (4–6)	600	300	52	3
TE (7–9)	600	300	52	3
TE (10–12)	600	300	52	3
TE (13–15)	600	300	52	3
TE (16–18)	600	300	52	3
TE (19–21)	600	300	52	3
Wine	148	30	14	3

Step 4. Select the number of the relative principle element based on the accumulative contribution rate:

$$\frac{\sum_1^i \lambda_i}{\sum_1^n \lambda_i} > 85\%. \quad (11)$$

Step 5. Use k NN for classifying.

5. Application Experiments

In order to verify the universal applicability and effectiveness of the new method, firstly, we test it by thirteen experiments based on Tennessee Eastman process dataset and the Wine dataset [16] which are different in numbers of samples N , dimensions D , and classes C ; in addition, we divide all the classes in three or four classes in sequence to avoid the deliberate choosing; secondly, experiments are also performed on original PCA and traditional DPPCA, respectively. The datasets descriptions and experimental settings are shown in Table 1.

5.1. TE Process Example. We use the Tennessee Eastman process data as the testing samples which are obtained from the document. The dataset includes twenty-one classes and each class in the experiment has two hundred train samples and one hundred test samples; besides, each sample has 52 features. The different types in TE process are overlapped and difficult to classify in observed space.

After the data preprocessing, the number of the principle elements is determined by the sum of elements being more than 85%. For the twelve experiments based on Tennessee Eastman process, we choose one of the experiments based on TE classes 5, 6, 7, and 8 labeled by F1, F2, F3, and F4, respectively, for detail introduction and demonstration. Its relative transformation operator for each feature to the TE process can be seen in Table 2 and the experimental results with the three approaches are shown in Figures 2(a)–2(c).

As we can see from the figures, PCA and DSPCA have poor performance in pattern classification where four classes are overlapped. In Table 4, their classification accuracy is 31.5% and 75.3%, respectively. Compared with them, the information entropy RPCA can help identify each class higher and its classification accuracy is 82.3%. Clearly, the InEnRPCA is a suitable feature extraction and classification method for fault diagnosis.

5.2. Wine Dataset. We select the Wine data from the machine learning database UCI, which includes three classes with the number of the each class being 59, 61, and 58; besides, each of the samples has fourteen features. The types in Wine are overlapped and difficult to classify in observed space.

The relative transformation operator for each feature in the Wine dataset can be seen in Table 3 and the experimental results with the three approaches are shown in Figures 3(a)–3(c).

From Figure 2(b) and Table 2, we can see that PCA and DSPCA do not have a perfect performance in pattern classification with the accuracy rate is 66.7% and 96.7%, respectively. Compared with them, the InEnRPCA can identify each class accurately, and the testing data can also be classified obviously, and its classification accuracy rate is 100%.

5.3. Discussion. From the experiments above, it is obvious that the proposed methods can successfully distinguish different fault types and implement fault pattern recognition and diagnosis in most times, via utilizing information entropy to calculate relative transformation operator for each feature and combining with relative PCA approach.

Meanwhile, we also find that although the proposed method can improve the fault classification performance in most times (nearly 80%), the method is sometimes inferior to DSPCA; the reason may be due to the overfitting for the train set and the dissimilarity between the train set and test set; however, we can make sure that utilizing the information

TABLE 2: Weight for each feature in the TE dataset.

Feature	Weight
1	1.26
2	0.83
3	1.16
4	1.06
5	1.15
6	0.82
7	1.25
8	0.94
9	0.89
10	1.29
11	1.25
12	0.96
13	1.26
14	1.12
15	0.90
16	1.23
17	0.80
18	1.06
19	1.37
20	1.35
21	1.14
22	1.23
23	1.27
24	0.78
25	1.25
26	0.58
27	0.67
28	1.27
29	1.32
30	0.94
31	1.30
32	0.54
33	0.67
34	1.31
35	1.23
36	1.23
37	0.59
38	1.30
39	0.60
40	0.65
41	0.78
42	1.17
43	1.25
44	1.26
45	1.53
46	1.34
47	1.31
48	0.96

TABLE 2: Continued.

Feature	Weight
49	0.90
50	1.37
51	1.22
52	1.41

TABLE 3: Weight for each feature in the Wine dataset.

Feature	Weight
1	0.53
2	0.28
3	0.13
4	0.28
5	0.20
6	0.42
7	0.71
8	0.17
9	0.28
10	0.59
11	0.43
12	0.57
13	0.59
14	1.56

TABLE 4: The classifying accuracy rate for each method.

Dataset	PCA (%)	DSPCA (%)	InEnRPCA (%)
TE (1-4)	50.5	95.8	97.3
TE (5-8)	31.5	75.3	82.3
TE (9-12)	34.0	41.8	47.5
TE (13-16)	38.3	54.5	53.8
TE (17-20)	47.3	77.0	78.0
TE (1-3)	68.7	100.0	100.0
TE (4-6)	62.0	64.0	59.0
TE (7-9)	47.0	73.3	77.3
TE (10-12)	45.7	58.7	59.3
TE (13-15)	49.0	59.3	38.3
TE (16-18)	62.0	83.3	85.0
TE (19-21)	32.7	55.3	59.3
Wine	66.7	96.7	100

entropy to calculate the weight for each feature to avoid distribution equally between different features is reasonable; we will continue in-depth research for better fault diagnosis.

6. Conclusion

This paper has analyzed the PCA method in the perspective of information theory and proposed a kind of fault diagnosis research based on information entropy and Relative Principle

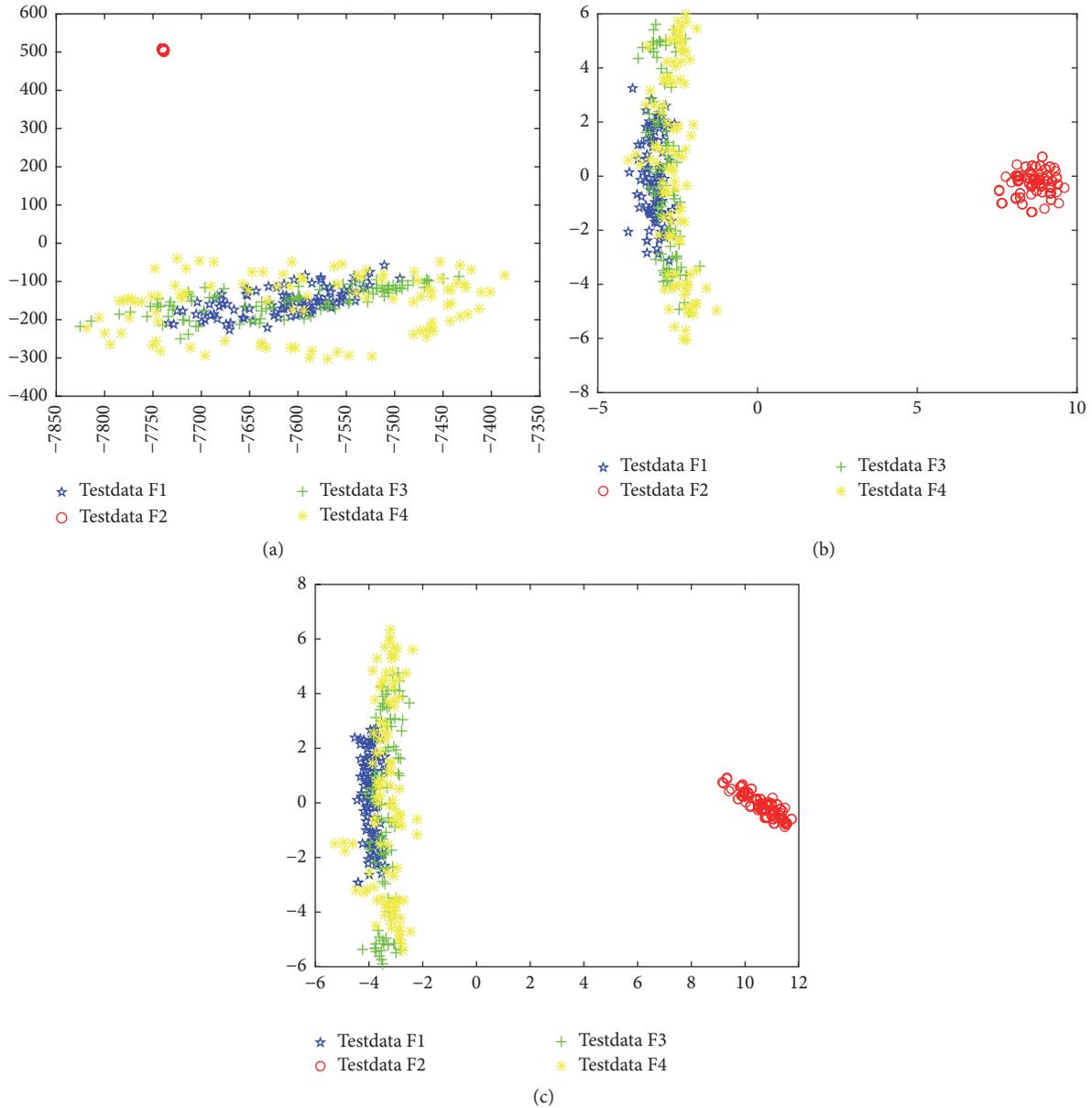


FIGURE 2: Feature extraction to TE process: (a) PCA. (b) DSPCA. (c) InEnRPCA.

Component Analysis. This proposed method can solve the following problems effectively: (1) the different result only due to calculating in different perspective of dimensions; (2) being uniformly distributed between the characteristic variables after dimension standardization; (3) how to calculate the weight for each variable. Both theoretical analysis and simulation experiments on Tennessee Eastman process and the Wine dataset from UCI demonstrate the feasibility and effectiveness of the new approach.

It is worth noting that the idea of using the information entropy to determine the certainty of the features is not new. However, the idea of combining information entropy with RPCA for faulty detection, to our knowledge, should have no previous publication. The proposed method should be seen

as an alternative fault diagnosis method. It is not superior to the other methods in all cases.

There is some interesting future work:

- (i) To exploit the way to make the energy conservation the same before and after the relative transformation: in the perspective of the energy conservation, the system energy before and after the relative transformation may usually not be the same. If we can consider the energy conservation during the relative transformation, the effect may be much better.
- (ii) To apply the relative transformation by information entropy as data preprocessing in other methods: the proposed method is not limited to RPCA method; it

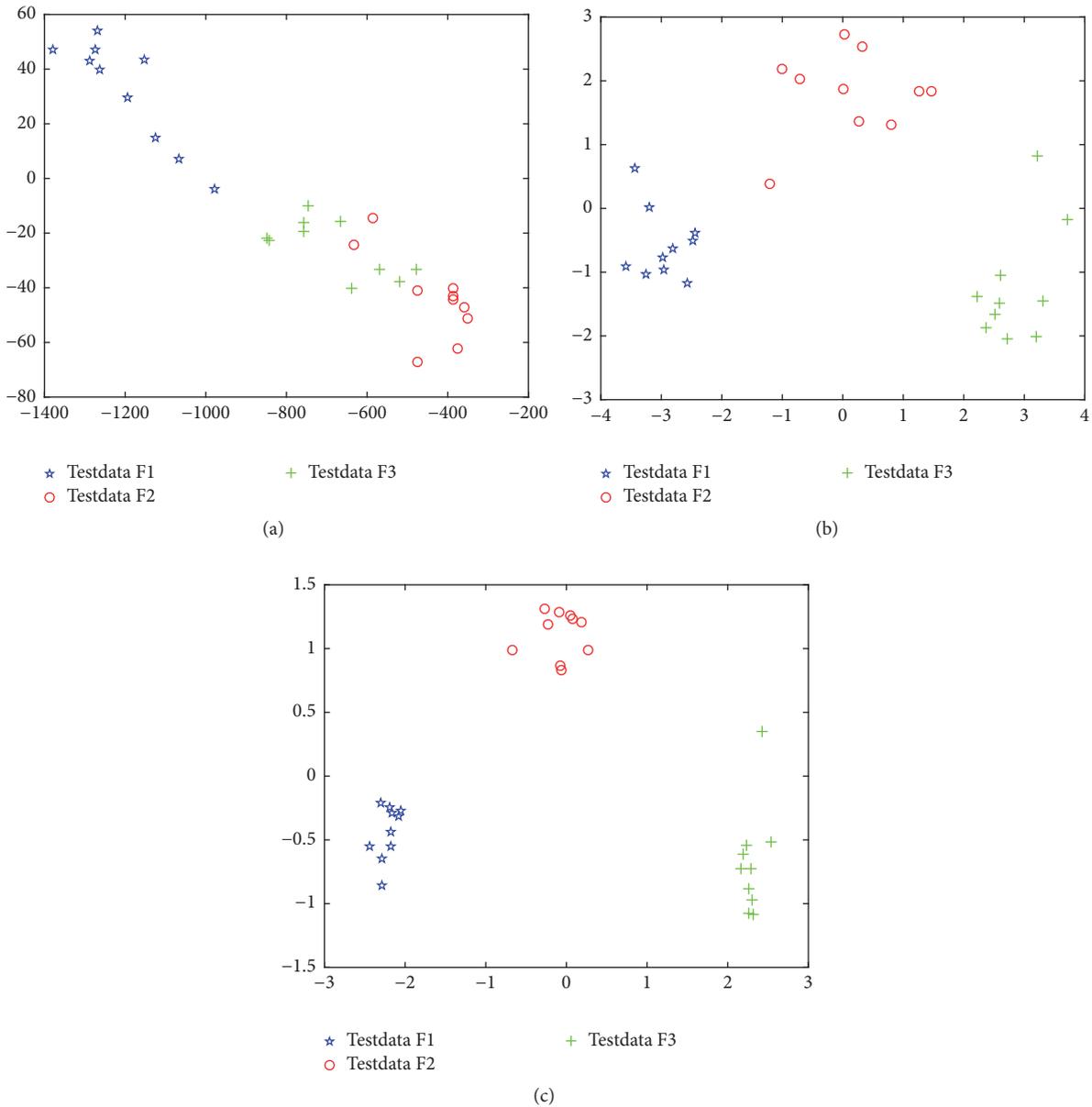


FIGURE 3: Feature extraction to Wine dataset: (a) PCA. (b) DSPCA. (c) InEnRPCA.

is an effective data preprocessing way that can also be applied to other fault diagnosis methods.

Competing Interests

The authors declare that they have no competing interests.

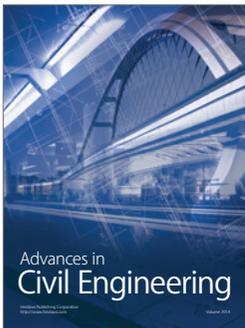
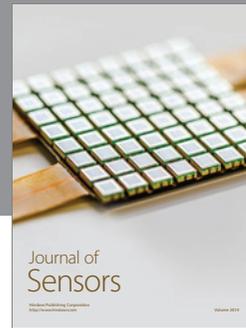
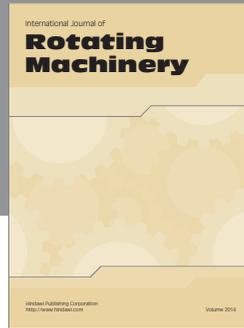
Acknowledgments

This work is supported by National Natural Science Foundation (NNSF) of China under Grant nos. U1509203, 61333005, U1664264, and 61490701.

References

- [1] N.-Y. Lu, F.-L. Wang, F.-R. Gao, and S. Wang, "Statistical modeling and online monitoring for batch processes," *Acta Automatica Sinica*, vol. 32, no. 3, pp. 400–410, 2006.
- [2] T. Z. Wang, *Intelligent Integration of Data Mining Method and Its Application*, Shanghai Maritime University, 2006.
- [3] G. H. Wen, "Relative transformation for machine learning," *Journal of Computer Research and Development*, vol. 45, no. 4, pp. 612–618, 2008.
- [4] H. T. Shi, J. C. Liu, and Y. W. Zhang, "Principle analysis faulty diagnosis based on relative transformation," *Journal of Huazhong University of Science and Technology*, vol. 37, supplement 1, pp. 70–73, 2009.

- [5] Y. Tang, T. Peng, Y. Xiong, and F. Jiang, "Breakdown voltage prediction method for transformer oil based on relative transformation principal component analysis," *Yi Qi Yi Biao Xue Bao/Chinese Journal of Scientific Instrument*, vol. 36, no. 7, pp. 1640–1645, 2015.
- [6] J. Yi, D. Huang, S. Fu, H. He, and T. Li, "Optimized relative transformation matrix using bacterial foraging algorithm for process fault detection," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 4, pp. 2595–2605, 2016.
- [7] W. Cheng-Lin, H. Jing, W. Tian-Zhen, and C. Zhi-Guo, "Relative PCA with applications of data compression and fault diagnosis," *Acta Automatica Sinica*, vol. 34, no. 9, pp. 1128–1139, 2008.
- [8] J. Hu, *Relative PCA and Its Application Research*, Henan University, 2008.
- [9] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.
- [10] Q. Hu, X. Che, L. Zhang, D. Zhang, M. Guo, and D. Yu, "Rank entropy-based decision trees for monotonic classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 11, pp. 2052–2064, 2012.
- [11] Y. Y. Chen and Y. M. Chen, "Attribute reduction algorithm based on information entropy and ant colony optimization," *Journal of Chinese Computer System*, vol. 36, no. 3, pp. 586–590, 2015.
- [12] X. L. Wang, L. Y. Lu, and W. P. Tai, "Research of a new algorithm of words similarity based on information entropy," *Computer Technology and Development*, vol. 25, no. 9, pp. 119–122, 2015.
- [13] Q. S. Meng, *Information Theory*, Xi'an Jiaotong University Press, 1986.
- [14] H. Li, *Statistical Learning Methods*, Tsinghua University Press, 2012.
- [15] T. Wang, Y. Liu, T. Tang, and Y. Chen, "Dynamic data window fault detection method based on relative principal component analysis," *Transactions of China Electrotechnical Society*, vol. 28, no. 1, pp. 142–148, 2013.
- [16] C. Blake, "UCI repository of machine learning databases," in *Neural Information Processing Systems*, 1998.



Hindawi

Submit your manuscripts at
<https://www.hindawi.com>

