Neurocomputing 190 (2016) 117–123

Contents lists available at ScienceDirect

Neurocomputing

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

Fed-batch fermentation penicillin process fault diagnosis and detection based on support vector machine

Chengming Yang [a,*], Jian Hou [b]

[a] Research Institute of Intelligent Control and Systems, Harbin Institute of Technology, Heilongjiang 150001, China
[b] College of Engineering, Bohai University, Liaoning 121013, China

ARTICLE INFO

Article history:
Received 11 October 2015
Received in revised form
7 December 2015
Accepted 6 January 2016
Communicated by Xudong Zhao
Available online 2 February 2016

Keywords:
Fault detection
Fault diagnosis
Support vector machine
Principal components analysis
Fed-batch fermentation penicillin

ABSTRACT

With the increase of scale and complexity of modern chemical process, fault diagnosis and detection are playing crucial roles in process monitoring. Accidents can be avoided if faults can be detected and excluded in time. In this paper, Principal Components Analysis (PCA) and Recursive Feature Elimination (RFE) are combined with Support Vector Machine (SVM) for fault diagnosis and detection. Specifically, the original SVM, PCA-SVM and SVM-RFE are respectively utilized to identify three faults from the simulation of Fed-Batch Fermentation Penicillin (FBFP) process. Experimental results show that PCA-SVM and SVM-RFE perform better than the original SVM, and the fault detection schemes based on PCA-SVM and SVM-RFE generate satisfactory results.

1. Introduction

Industrial production systems have become more and more diverse and complex with the development of industrial technology [1–4]. As a result, it is critical to apply fault detection and diagnosis techniques to the key variables in modern industrial processes [5–8]. Fault detection and fault diagnosis technology is widely used in monitoring light fermentation [9,10], chemical processes [11–13] and other industrial processes [14,15], and especially multivariate statistical process control (MSPC) method has become a big concern in industry and academia recently [16–18]. Data will be mapped onto a lower dimensional space from a high dimensional space based on MSPC method [19], in order to reduce the interference information and keep the characteristics of the initial data. The commonly used methods, e.g., principal components analysis (PCA) [20], correspondence analysis (CA) [21], principal components regression (PCR) [22], canonical variate analysis (CVA) [23], partial least square (PLS) [24], and independent component analysis (ICA) [25], have been widely used in industrial processes. The work of this paper is focused on the fault detection and diagnosis of the fed-batch fermentation penicillin (FBFP) process which has been extensively studied in statistical process monitoring and process control.

Support Vector Machine (SVM) as a machine learning method based on statistical theory is much suitable for small sample data. After it was proposed, it caused an extensive attention of people to research in this field, since its excellent learning performance especially the generalization ability [26–28]. In this paper, SVM and PCA are introduced in the first place. Since SVM has a sound theoretical foundation and remarkable generalization ability, it has been widely applied in pattern recognition, classification and regression problems. With PCA, the dimension of variables can be reduced and the information of the original data will be remained as much as possible. Recursive feature elimination (RFE) is used to sort the features in descending order. The PCA-SVM algorithm is outlined followed by SVM-RFE. Three kinds of faults are classified respectively by original SVM, PCA-SVM and SVM-RFE finally.

2. Related works

2.1. Support vector machine

According to statistical learning theory [29,30], SVM was proposed in 1995 as a binary classifier. SVM is designed based on the Vapnik–Chervonenkis (VC) Dimension theory and structural risk minimization (SRM) principle. Hence, the VC dimension can be seen as the best indicator of the learning ability of function set. A high VC dimension means a high complexity of the problem and a low generalization ability of learning methods. Since SVM is designed based on the VC dimension, its classification result is

* Corresponding author.
E-mail addresses: chmyang@foxmail.com (C. Yang), jian_hou@163.com (J. Hou).

http://dx.doi.org/10.1016/j.neucom.2016.01.027
0925-2312/

independent of sample dimension. The SRM theory has been proposed to minimize the structural risk, i.e., the sum of empirical risk and confidence risk. By means of SRM, SVM possesses remarkable generalization ability.

The original SVM can be applied to two-class classification problems. The 1-v-r SVMs, 1-v-1 SVMs and H-SVMs methods can be used to solve multi-class classification generally.

In the 1-v-r SVMs method, each class of samples are regarded as in a category, and the rest are treated as in another category. In total c classification models will be constructed for classification problem with c class of samples. The advantages of 1-v-r SVMs lie in the simplicity and computation efficiency. However, the shortcoming is that the classification may be inseparable. If there are many classes of samples, and one class of the samples is considerably less than the sum of the others, this imbalance will affect the classification accuracy.

In the 1-v-1 SVMs method, the classification models are trained with every two classes of samples. In total $\frac{c(c-1)}{2}$ classification models are trained for c classes of training samples. The training samples of each classifiers are relevant, and the voting method is used for classification.

All classes of samples are divided into two sub-categories in H-SVMs method, and the sub-categories will be divided into another two categories after tested by SVM. After some iterations, the single category can be obtained. Therefore H-SVMs method avoids the case that samples cannot be divided, and it has a stable promotion performance [31].

2.2. Principal components analysis

PCA was firstly proposed in 1901. Mudholkar and Jackson [32] introduced PCA into multivariate statistical process control and brought the squared prediction error (SPE) into PCA. Noise-sensitive characteristic of residual space was utilized to monitor the inapparent data points in T^2 statistics, in order to increase the accuracy of fault diagnosis and detection. As a multivariate statistical method, PCA is studied by academia and industry extensively. Multiple related variables of data sets are converted into a few irrelevant variables based on the PCA statistics. It can not only reduce the dimension of variables, but also keep the information of the original data as much as possible. According to the PCA statistics, if the index of test data exceeds the threshold calculated from normal data, the test data will be treated as faulty [33,12].

3. Algorithm

3.1. The original SVM

SVM shows some unique advantages over other approaches in classification with linearly inseparable data, small samples and high feature dimension. It deals with these cases by means of error penalty, slack variables and kernel functions.

With the original SVM, the input data should belong to two categories. The labels of the data in the positive category are $y_i=+1$, and the labels of those in the negative category are $y_i=-1$. The training samples can be written as (x_i, y_i) $(i=1,2,\ldots,l)$, $x \in R^n$.

The separating hyperplane is

$$(w \cdot x)+d=0 \tag{1}$$

where w is the weight vector and d is a constant. To ensure that all samples are correctly classified and have the category interval, the following condition should be satisfied

$$y_i[(w \cdot x_i)+d] \geq 1 \tag{2}$$

The separating interval can be calculated as $\frac{2}{\|w\|}$. The objective function to obtain the separating hyperplane can be written as

$$\min \phi(w)=\frac{1}{2}\|w\|^2 \tag{3}$$

In order to solve Eq. (3), Lagrange function is introduced

$$L(w,d,a)=\frac{1}{2}\|w\|-a(y((w \cdot x)+d)-1) \tag{4}$$

In Eq. (4), $a(a_i>0)$ is the Lagrange multiplier. Then the problem is converted to the corresponding dual problem

$$\max\ Q(a)=\sum_{j=1}^{l} a_j-\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} a_i a_j y_i y_j (x_i \cdot x_j)$$

$$\text{s.t.} \sum_{j=1}^{l} a_j y_j=0,\ j=1,2,\ldots,l,\ \ a_j \geq 0,\ j=1,2,\ldots,l \tag{5}$$

The optimal solution calculated by Eq. (5) is $a^*=(a_1^*, a_2^*, \ldots, a_l^*)^T$, and the optimal weight vector w^* and d^* are written as

$$w^*=\sum_{j=1}^{l} a_j^* y_j x_j \tag{6}$$

$$d^*=y_i-\sum_{j=1}^{l} y_j a_j^* (x_j \cdot x_i) \tag{7}$$

Therefore, the optimal separating hyperplane is written as $(w^* \cdot x)+d^*=0$, and the optimal separating function is

$$f(x)=\operatorname{sgn}\{(w^* \cdot x)+d^*\}=\operatorname{sgn}\left\{\left(\sum_{j=1}^{l} a_j^* y_j (x_j \cdot x_i)\right)+d^*\right\} \tag{8}$$

To solve the nonlinear problems, SVM maps the low dimensional vectors into a higher space using kernel function according to nonlinear transform

$$x \to \phi(x)=(\phi_1(x), \phi_2(x), \ldots \phi_l(x))^T \tag{9}$$

By replacing the input vector x with eigenvector $\phi(x)$, the optimal separating function can be written as

$$f(x)=\operatorname{sgn}(w \cdot \phi(x)+d)=\operatorname{sgn}\left(\sum_{i=1}^{l} a_i y_i \phi(x_i) \cdot \phi(x)+d\right) \tag{10}$$

3.2. SVM-RFE algorithm

RFE aims at sorting the features in descending order. The effectiveness of feature selection can be improved by combining RFE with SVM [34]. The feature ranking list is constructed by SVM-RFE according to the weight vector w in the linear classifier. The classical SVM-RFE uses the linear kernel function, while RBF kernel function is introduced into SVM in non-linear cases [35].

In each iteration, the most irrelevant feature will be removed. The feature removed first is the last one in the list, and this means the feature is unnecessary. According to the feature ranking list, the most relevant variable of the faults can be obtained. Hence, the classification accuracy can be improved by the optimal subset consisting of the relevant features. The first feature in the list is treated as the most relevant one to be analyzed [36].

3.3. PCA-SVM algorithm

PCA is a classical data dimension reduction method. Redundant information in the original data space can be excluded by the principal component. Meanwhile, a lot of variance information is retained and each principal component variable is mutually orthogonal [37].

Let us assume that $X \in R^{n \times m}$ is a data matrix under normal conditions, and m is the number of variables, n represents the