ISSN : 0974 - 7435 Volume 10 Issue 11

2014

BioTechnology

An Indian Journal

FULL PAPER

BTAIJ, 10(11), 2014 [5402-5407]

Application of support vector machine in fault diagnosis based on PCA

Yan-wei Zhu

Department of Mathematics and Information science, Tangshan Normal University, Tangshan 063000, (CHINA)

ABSTRACT

Fault diagnosis is an important research subject in modern production. The production process of abnormal will bring great losses to enterprises. Therefore it is important to adopt effective fault diagnosis of the production process. SVM will be use to industrial process fault diagnosis. The advantages of the two methods will be fully reflected. The example proves the feasibility and efficiency of this method.

KEYWORDS

Support vector machine; PCA; Fault diagnosis; Kernel function.

INTRODUCTION

In recent years, with the development of modernization, the need for fault diagnosis technology is gradually enhanced. The research on the fault diagnosis method of China rapid advance, in practical applications the safety and reliability have become a key factor of economic benefit and social benefit. It received extensive attention. From the development of the discipline theory, Fault diagnosis has a strong interdisciplinary nature.

Modern control theory, Signal processing, pattern recognition, optimization, decision, etc are booming in recent 20 years. These are the theoretical foundation to solve the fault diagnosis problem of complex industrial systems. In order to avoid the occurrence of the fault in the process of production, countries have invested a lot of manpower to fault mechanism and fault diagnosis technology.

Due to the continuous expansion of the scale of industrial process, the process is complex, and the huge investment, but also forces people to consider there liability and safety of the process. The existing research results confirmed, technical process and engineering system changes many process or system failure related problems and can be converted into with this kind of problem[6-9]. Therefore the research is extremely relevant technology to detect process faults, there are profound theory value and can not be ignored the importance in the modern production process. In this paper support vector machine based on PCA is used to fault diagnosis. It plays algorithm advantages, also brings certain economic benefits.

THE PRINCIPLE OF SUPPORT VECTOR MACHINE

Support vector machine (SVM)[1,4] is a general learning algorithm to solve the small sample learning problem; this method was proposed by AT&TBell Vapnik et Laboratory. It was originated from the statistical learning theory and the development of the theory of VC. To the mid-90s, this theory has been progressively developed shape and were widespread concerned. The structural risk minimization (SRM) principle with the algorithm, SVM achieved SRM principle. Among the many classifiers, linear classifier has the simplest structure, people will consider the method with linear discriminate function similar to implement SRM principle, SVM is proposed from the model of linear optimal classification in case of surface (Optimal Hyper plane) and its basic idea is: if the structure of the classifier in the original feature space is a very complex, through defining the appropriate kernel function to induce a nonlinear transform, this transform the input vector X is mapped in the original feature space to a high dimensional feature space Z, and then in the new feature space to construct the optimal Hyper plane, it will reduce the complexity of the classifier. (as shown in Figure 1).

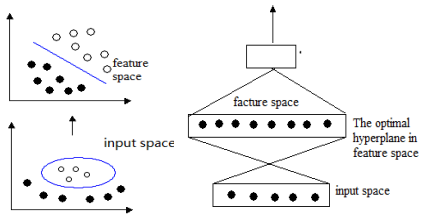

Figure 1 : The mapping between the input space and the feature space

When the kernel function is selected to meet certain conditions, by the two characteristics of high dimensional feature space of the kernel function is derived in vector dot product of the kernel function in the low dimensional feature space defined on the feature vectors corresponding to two and. In this way, we can deal with the high dimensional feature space of the data in the low dimensional feature space. The solution of the SVM refers only to the vector dot product computation, so we don't have to worry about the calculation of dimension disaster caused by introducing the kernel function, and can

concentrate on how to select the appropriate function, in order to improve the feature vectors in the high dimensional feature space distribution, so that the classifier structure is more simple. So, the solving process of SVM is in the high dimensional feature space for optimal classification pattern sample data process, optimal classification face here is the classification interval between two kinds of sample data in the premise of control sample misclassification rate classification.

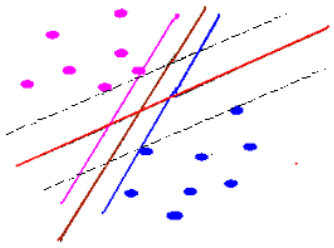

Figure 2 : The principle of SVM in two dimensional space

In Figure 2 two dotted lines between the red lines is the SVM which corresponds to the optimal solution of the corresponding classifier, map and two dotted line tangent point is support vector which corresponds to the optimal solution with non-zero sample. The distance between two lines two times Δ interval. SVM corresponds to the maximum gap. Solid line in the figure intervals than optimal interval should be small, among them the corresponding classifier can also sample data without separate errors, but the classifier's generalization ability is inferior to the SVM.

d dimensional space in the general form of linear discriminated function as follows:

$$w \cdot x + b = 0 \quad \textbf{(1)}$$

The optimal separating line is not only be able to separate the two groups (training error rate is 0), but also the largest class interval, on H_1 and H_2, the training sample points called support vector.

Linear separable sample set as follows:

$$(x_i, y_i), i = 1, \cdots, n,\ x \in R^d,\ y \in \{+1, -1\}$$

Lagrange optimization method can be obtained using the optimal classification surface of the optimal classification function is:

$$f(x) = \text{sgn}(\sum_{i=1}^{n} y_i \alpha_i^* (x_i \cdot x) + b^*) \quad \textbf{(2)}$$

Where : α_i^* is optimal solution of Lagrange multiplies α_i, b^* is the classification threshold. α_i of the non-support vector is 0, Therefore, where the sum is only a matter of support vectors. If dot product is instead by inner product $K(x_i, x)$, that is a certain feature space transformed into a new feature space, the corresponding discriminated function as follows:

$$f(x) = \text{sgn}(\sum_{i=1}^{n} y_i \alpha_i^* K(x_i, x) + b^*) \quad \textbf{(3)}$$

The principal component analysis overview

In dealing with multi-sample data, the first problems encountered were a lot of observation data. As a whole

$X = (x_1, \cdots, x_p)'$,

X is One-dimensional random vector of p,where $x_i \; i = 1, \cdots, p$, if there are n sample : $X_1, X_2, \cdots, X_n$,there are $n \times p$ data. How to tackle the main laws of these data to analyze samples or general primary in nature? If multiple indicators are independent of each other, then the problem can be dealt into p single indicator, which is simple and unusual situation. Principal component analysis method is proposed, it is a multi-index into a few composite indicator of a statistical analysis method. In practical problems, multi-index problem is frequently encountered. Multi-variable (multi-index) problems are studied in multivariate statistical analysis. The mechanism can be simply stated as follows: T is a orthogonal transformation, the original random variables $X = (x_1, \cdots, x_p)'$, which are related, turn into a new random vector

$U = (u_1, u_2, \cdots, u_p)'$

They are not related. In the algebra showed the X covariance matrix will be a pair angular transformation matrix. They reflected in the geometry which transforms the original coordinate system into a new orthogonal coordinate system, and point to sample point spread of the most open p orthogonal direction. Then multi-dimensional variables are treated by low-dimensional. So that it is a high accuracy of low-dimensional variables into the system[2,3].

THE CONTENT OF INDUSTRIAL PROCESS FAULT DIAGNOSIS

The main object of study process monitoring is non congenital failure process runtime changes or system components[9,10].

(1) Process changes

Process of abnormal changes mainly refers to object of the state or itself a characteristic change, it usually refers to the process state deviation from the normal state, such as: abnormal changes of various instantaneous appearing in the running process of the system or continuous, slow or sudden, partial or whole body, exceeding one's expectations or otherwise not expected is usually divided into two types: gradual change and mutation. According to the anomaly duration, mutation can be further divided into pulse type step mutation, mutation and compound mutation types such as.

(2)System failure

System failure can be understood as there is at least one important variable or characteristic deviates from the normal range. Generally speaking, system failure can be understood as any abnormal phenomena of the system, The characteristics of the system exhibited undesirable, fault in at least two different meanings: one refers to the defect or flaw system function components of innate; one is fault dynamic system functional system in use or abnormal operation of the change as the main object of process monitoring, mostly failure process is dynamic system in this paper refers to a class of. In general, the dynamic system of fault location time characteristic form of diversity. According to the different positions can be divided into components fault: that certain components in object, or even abnormal subsystem, which makes the whole system, can not be normally complete its function.

EXPERIMENTAL RESULTS

This paper selects the Tennessee (TE) typical fault in the process. TE process is a real process simulation of a chemical process, it is a standard test procedure proposed by JJ.Downs and E.F.Vogel in

1993. The TE process data downloaded from the http://brahms.scs.uiuc.edu.[11] There are five main operating units in TE production process:

There are many factors such as: reactor, gas / liquid separator, compressor, product decoration tower. In this paper the simulation experiment was carried out by simulation software MATLAB6.5[5]. The typical faults are 4, 9, 11. Because the 4 fault and fault 11 were related to the reactor cooling water entrance temperature, the fault is different type. Fault 4 is a step change, but fault 11 is random variation. When two kinds of failure happened, there is overlap area between the deviations from the normal variables. 9 is the fault of random changes, and it is about D. it is different from Fault 4 and 11. Interleaved data sets are three kinds of faults. Each group has two groups of fault data, the training data used to build the model, test data are used to validate the model. The training data contains 480 groups of data; test data contains 960 data sets. Each observation value contains 52 process variables. Only 9 variables (reactor temperature) and variable 51 (reactor cooling water flow) are very important to distinguish between these three kinds of fault. Because the 4 fault and 11 are related to the reactor, while Fault 9 is linked to the D feed temperature. And it will influence the reactor. The other 50 variables had no significant effect for the three kinds of fault. Fault 4 and fault 9 can easily be distinguished according to these two variables. In the classifier, RBF kernel function, linear kernel function, Poly kernel function is selected. Comparing the results shown in the table below

TABLE 1 : The original data of diagnosis results

	Correct rate /%	Test time /s
RBF Kernel	78.65	16.23
Linear Kernel	76.51	18.32
Poly Kernel	77.98	20.21

TABLE 2 : Results of fault diagnosis SVM based on PCA

	Correct rate /%	Test time /s
RBF Kernel	87.37	13.55
Linear Kernel	83.62	16.51
Poly Kernel	85.63	17.35

SUMMARY

The support vector machine algorithm can make full use of the original data information to fault classification. In this paper, the concept of principal component analysis is added to the algorithm, fully embodies the advantages of support vector machine algorithm. The test results also indicated that the support vector machine based on principal component analysis in both computing time and in the correct rate of classification has great superiority. At the same time, the support vector machine because of its own characteristics, suitable for processing the ill posed problem, defined over fitting, non-linear data sets such as effective treatment, but also has great potential in engineering application; this study will be an important development trend of support vector machine.

ACKNOWLEDGMENT

This work is supported by Hebei Province Natural Science Fund Project (No. A2012105021), technology to guide the project of higher education of Hebei Province (No.Z2011328) and the key project of natural science of Tangshan Normal University (2014D06).

REFERENCE

[1] V.Vapnik write, Zhangxuegong translation; The Nature of Statistical Learning theory[M]. Beijing: Tsinghua University Press, **(2000)**.
[2] Gao huixuan; Applied multivariate statistical analysis. Beijing Beijing University Press, **(2005)**.
[3] Qin shoukang; Principle and application of comprehensive evaluation. Beijing Electronics Industry Press.
[4] Deng naiyang, Tian yingjie; A New Method in Data Mining : Support Vector Machine. Beijing : Science Press, **(2004)**.
[5] Weiwei; MATLAB: Application of mathematical toolboxTechnical Manual. Beijing, National Defense Industry Press, **(2004)**.
[6] Liu peijie, Huangqiang, Cheng yuqiang, Wu jianjun; Liquid rocket engine fault diagnosis method of dynamic cloudbased on BP network. Journal of Aerospace Power, **12**, **(2012)**.
[7] Chen Wenjie, Wangjing; Application of support vector machine in industrial process. Computers and Applied Chemistry, **22(3)**, 3 **22(3)**.
[8] Xu Qiaoling, Lin Weihao, Zhaochao, Huang Yunyun; Industrial boiler flue gas oxygen content prediction based onKPCA-LS-SVM. Computers and Applied Chemistry, **29(7)**, 7 **(2012)**.
[9] Liu Xiliang, Chen Guiming, Li Fangxi, Zhangqian; Research and Application of Bayesian Network Fault Diagnosis Method Based on Multi-Sensor Information Fusion. Mechanical Science and Technology, **1**, **(2013)**.
[10] Li Jianhong, Jiang Tongmin, He Yuzhu, Jiang Jueyi; SVM fault diagnosis method based on NMF. Journal of Beijing University of Aeronautics and Astronautics, **12**, **(2012)**.
[11] Li Yuheng, Zhaofeng; Application of Support Vector Machine in Fault Diagnosis of Transformer, Technology Innovation and Application, **16**, **(2014)**.