

TWO-LEVEL STRUCTURE FOR ANALOG CIRCUIT FAULT DIAGNOSIS USING IMPROVED DAGSVM AS CLASSIFIER

KE GUO, YI ZHU, YE SAN

School of Astronautics, Harbin Institute of Technology, Harbin 150001, Heilongjiang, China

ABSTRACT

Fault diagnosis of analog circuit is of essential importance for guaranteeing the reliability and maintainability of electronic systems. Taking into account the requirements and characteristics of analog circuit fault diagnosis, two-level diagnostic structure for analog circuit is proposed in this paper. Analog circuit fault diagnosis can be regarded as a pattern recognition issue and addressed by multi-class SVM. Aiming at the uncertainty of the node arrangement and the error accumulation phenomenon, the improved directed acyclic graph support vector machine (DAGSVM) based on fisher separability measure in high dimensional feature space and margin of SVM is proposed. In order to eliminate redundant fault features and the impact of measure noise on the diagnostic accuracy, simultaneously taking into account lightening the workload of the classifier, the separability promotion algorithm based on kernel principal component analysis (KPCA) plus linear discriminant analysis (LDA) is proposed. Then the separability promotion algorithm is used to extract the nonlinear fault features of analog circuit and the two-level structure based on improved DAGSVM is applied to diagnose faults in analog circuit. The unbalanced classification model based on SVM is adopted in the first level. The effectiveness of the proposed method is verified by the experimental results.

Keywords: *Fault Diagnosis, Analog Circuit, Improved DAGSVM, Separability Promotion Algorithm, Kernel Principal Component Analysis*

1. INTRODUCTION

The normal operations of aircraft are guaranteed by the reliable avionic equipment to a considerable extent. With the integration and the increasing scale of electronic devices, the reliability and research of avionics equipment are of great significance [1]. Generally speaking, the circuit system of avionic equipment can be divided into digital circuit and analog circuit. Above 80% of the circuits in electronic equipment is digital, but virtually 80% of the faults occur in analog circuits [2, 3]. Fault diagnosis of digital circuit has reached the point of automation, while that of analog circuits is still challenging due to component tolerances, circuit nonlinearities measurement noise and poor fault models [4]. These difficulties make the application of artificial intelligence techniques to these problems very attractive. Analog circuit fault diagnosis can be regarded as a pattern recognition issue and addressed by artificial neural network or support vector machine. Neural Network has excellent classification capability and self-learning ability, and can approximate any nonlinear function in theory when the number of hidden layer nodes is sufficient.

Recently, applications of neural networks to analog fault diagnosis are studied comprehensively and exhaustively [5-8]. In literature [5], the BP neural network is adopted to the fault diagnosis of linear circuits. Because there is no preprocessing for the impulse response of the circuit, even for a relatively small circuit, complex neural networks architecture is demanded for this approach. In order to solve this problem, in literature [6], the authors have applied BP neural network with wavelet decomposition and principal component analysis (PCA) as preprocessors to fault diagnosis of analog circuits. Compared with literature [5], the proposed approach demands a much smaller network and has better diagnostic accuracy. The development and application of neural networks has always been plagued by multiple local minima, proneness to over-fitting and the empirical nature of model construction, though it offers a number of advantages. Being different from neural networks, support vector machine (SVM) is built on the VC dimension theory and structural risk minimization principle, having the advantages of global and unique optimum, simple structure, the better classification ability and generalization performance. Based on these good characteristics, SVM is gradually proposed to diagnose analog

circuit in recent years [9-11]. In literature [9], a novel analog circuit fault diagnosis approach based on improved one-against-rest SVM is proposed, but the drawback of the approach is that the number of the required accessible nodes is up to 6. In literature [10], a new approach of fault diagnosis in analog circuits, which employs the fractional wavelet transform to extract fault features and adopts a fuzzy multi-classifier based on the support vector data description to diagnose circuit faults, is proposed. A threshold value is adopted to decrease the fuzzy region which in the overlap between hyperspheres, but it is difficult to select a reasonable threshold.

An improved directed acyclic graph SVM method is proposed and then adopted to diagnose analog circuit faults in this paper. Taking into account the importance of preprocessing the output data of the circuit under test, kernel principal component analysis (KPCA) plus linear discriminant analysis (LDA) is adopted as a preprocessor. The paper is organized as follows. In section 2, the principle and implementation steps of separability promotion algorithm based on kernel principal component analysis and linear discriminant analysis are given. In section 3, the improved DAGSVM is given. This is followed by two-level structure for analog circuit fault diagnosis and the experimental results. Finally, conclusions are given in section 5.

2. SEPARABILITY PROMOTION ALGORITHM BASED ON KPCA PLUS LDA

KPCA is first developed by Schölkopf [12] based on the theory of reproducing kernel Hilbert space. Due to better performance of nonlinear features extraction and noise elimination as compared with PCA, KPCA has gained substantial attention as a learning machine in pattern recognition, statistical analysis and image processing. The features extracted by KPCA are not optimal classification features but optimal description features. In order to solve the problem, a novel approach of data reconstruction by KPCA and processed by LDA which named separability promotion algorithm is proposed.

2.1 Kernel Principal Component Analysis

Consistent with Cover's theorem, it has been shown that nonlinearly separable patterns in the original input space may be transformed such that they are linearly separable in a high dimensional space via a nonlinear mapping. This high

dimensional linear space is referred to as the feature space F . The basic idea of KPCA is to map the input data $x \in R^m$ into a new high dimensional feature space F firstly via a nonlinear mapping $\phi(\cdot)$, and then perform a linear PCA in F .

Let $X = [x_1, x_2, \dots, x_n]$ ($x_i \in R^m, i = 1, 2, \dots, n$) be the observation set, where n is the sample number, m is the number of variables. By the nonlinear mapping $\phi: x \mapsto \phi(x) \in F^h$, the measured inputs are extended into the high dimensional feature space, where h is the dimension in feature space which is assumed to be a sufficiently large number. The sample covariance matrix in the feature space can be expressed by

$$C_\phi = \frac{1}{n} \sum_{i=1}^n (\phi(x_i) - m_\phi)(\phi(x_i) - m_\phi)^T \quad (1)$$

where $m_\phi = \frac{1}{n} \sum_{i=1}^n \phi(x_i)$ is the sample mean. We denote $\bar{\phi}(x_i) = \phi(x_i) - m_\phi$ as the centered feature space sample.

Then formula (1) can be expressed as

$$C_\phi = \frac{1}{n} \sum_{i=1}^n \bar{\phi}(x_i) \bar{\phi}(x_i)^T \quad (2)$$

For convenience, we assume that $\phi(x_i)$ have been centralized, where $i = 1, 2, \dots, n$.

The kernel principal component can be obtained by solving the eigenvalue problem in the feature space:

$$\lambda v = C_\phi v = \frac{1}{n} \sum_{i=1}^n (\bar{\phi}(x_i), v) \bar{\phi}(x_i)^T \quad (3)$$

where eigenvalues $\lambda \geq 0$ and eigenvectors $v \in F$.

It is easy to see that every eigenvector v of C_ϕ lies in the span of $\phi(x_1), \dots, \phi(x_n)$. Hence $\lambda v = C_\phi v$ is equivalent to

$$\lambda(\phi(x_k), v) = (\phi(x_k), C_\phi v) \quad (4)$$

where $k = 1, 2, \dots, n$ and there exist coefficients α_i , $i = 1, 2, \dots, n$. Such that

$$v = \sum_{i=1}^n \alpha_i \phi(x_i) \quad (5)$$

Combining equations (4) and (5), we get



$$\lambda \sum_{i=1}^n \alpha_i (\phi(x_k), \phi(x_i)) = \frac{1}{n} \sum_{i=1}^n \alpha_i (\phi(x_k), \sum_{j=1}^n \phi(x_j)) (\phi(x_j), \phi(x_i)) \quad (6)$$

Defining kernel matrix K with size $n \times n$ by $[K]_{ij} = (\phi(x_i), \phi(x_j))$, then its elements are determined by virtue of kernel tricks.

$$[K]_{ij} = (\phi(x_i), \phi(x_j)) = \kappa(x_i, x_j) \quad (7)$$

where $\kappa(x_i, x_j)$ is the calculation of the inner product of two vectors in feature space F with a kernel function. This reads

$$n\lambda K\alpha = K^2\alpha \quad (8)$$

where α denotes the column vector with entries $\alpha_1, \alpha_2, \dots, \alpha_n$. To get solutions of equation (8), we solve the eigenvalue problem, and it is equivalent to perform PCA in F .

$$n\lambda\alpha = K\alpha \quad (9)$$

Let $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ denote the eigenvalues of K , and $\alpha_1, \alpha_2, \dots, \alpha_n$ the corresponding complete set of eigenvectors. The dimensionality of the problem can be reduced by retaining only the first p eigenvectors. We normalize $\alpha_1, \alpha_2, \dots, \alpha_p$ by requiring that the corresponding vectors in feature space F be normalized, i.e., $(v_k, v_k) = 1$ for all $k = 1, 2, \dots, p$. According to equation (5) we get

$$\left(\sum_{i=1}^n \alpha_i^k \phi(x_i), \sum_{j=1}^n \alpha_j^k \phi(x_j)\right) = 1 \quad (10)$$

Further, we get

$$\left(\sum_{i=1}^n \alpha_i^k \phi(x_i), \sum_{j=1}^n \alpha_j^k \phi(x_j)\right) = \lambda_k (\alpha_k, \alpha_k) \quad (11)$$

Knowing the normalized vectors, the PCs t of a test vector x are then extracted by projecting $\phi(x)$ onto eigenvectors v_k in F , $t_k = (v^k, \phi(x)) = \sum_{i=1}^n \alpha_i^k \kappa(x_i, x)$, where $k = 1, 2, \dots, p$, p is the number of principal components.

In order to avoid performing the nonlinear mappings and computing dot products in the feature space F , kernel function of form $\kappa(x, y) = (\phi(x), \phi(y))$ is introduced. Mercer's theorem

guarantees the existence of a number of kernel functions. The most popular kernel function is

Gaussian kernel (Radial basis kernel)

$$\kappa(x, y) = \exp(-\|x - y\|^2 / 2\sigma^2) \quad (12)$$

where σ is a positive real number.

Before applying KPCA, the centered kernel matrix \bar{K} should be calculated by

$$\bar{K} = K - 1_n K - K 1_n + 1_n K 1_n \quad (13)$$

where $1_n = \frac{1}{n} \begin{bmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{bmatrix}$.

2.2 Linear Discriminant Analysis

KPCA can enhance the separability between different classes to a certain degree when used as preprocessor for nonlinear feature extraction. However, for a multi-class problem with serious overlapping in between-class distribution, rising dimension is difficult to enhance the class separability effectively. The reason is that the features extracted by KPCA are not optimal classification features but optimal description features. To address this issue, separability promotion algorithm based on KPCA plus LDA is proposed which can enhance class separability and reduce the within-class scatter in high dimensional space. Consequently, the purpose of improving the classification accuracy can be achieved.

The basic idea of LDA is to perform linear projection of sample data, maximize projected data in between-classes divergence and minimize it in within-class divergence [13]. The mathematical description of LDA for c -class problem is as follows. Assuming that there are sample points of c classes in d -dimensional space.

$$X^{(i)} = \begin{bmatrix} x_1^{(i)} \\ x_2^{(i)} \\ \vdots \\ x_{N_i}^{(i)} \end{bmatrix} \quad (14)$$

where $x_j^{(i)} \in R^d$, $i = 1, 2, \dots, c$, $j = 1, 2, \dots, N_i$, N_i is the number of i th class samples, the total number of samples is $N = \sum_{i=1}^c N_i$, the mean of i th class sample defined as

$$m_i = \frac{1}{N_i} \sum_{j=1}^{N_i} x_j^{(i)} \quad (15)$$

The total sample mean defined as



$$m = \frac{1}{N} \sum_{i=1}^c \sum_{j=1}^{N_i} x_j^{(i)} = \frac{1}{N} \sum_{i=1}^c N_i m_i \quad (16)$$

Within-class scatter matrix of *i*th class sample is defined as

$$S_w^{(i)} = \sum_{j=1}^{N_i} (x_j^{(i)} - m_i)(x_j^{(i)} - m_i)^T \quad (17)$$

The total within-class scatter matrix S_w defined as

$$S_w = \sum_{i=1}^c \sum_{j=1}^{N_i} (x_j^{(i)} - m_i)(x_j^{(i)} - m_i)^T \quad (18)$$

Between-class scatter matrix defined as

$$S_B = \sum_{i=1}^c N_i (m_i - m)(m_i - m)^T \quad (19)$$

Finding a direction w with the maximum generalized Rayleigh entropy $J(w)$

$$J(w) = \frac{w^T S_B w}{w^T S_w w} \quad (20)$$

The optimization problem can be easily proved by Lagrange dual method

$$\arg \max_w \frac{w^T S_B w}{w^T S_w w} \quad (21)$$

which is equivalent to solve the generalized eigenvalue problem

$$S_B w = \lambda S_w w \quad (22)$$

The directions w with the maximum separability of the different classes can be obtained by solving formula (22).

The projection of a d -dimensional sample data $x_j^{(i)}$ on the direction of the vector w is

$$y_j^{(i)} = w^T x_j^{(i)} \quad (23)$$

Steps of the separability promotion algorithm based on kernel principal component analysis and linear discriminant analysis are as follows.

The separability promotion algorithm based on KPCA+LDA

- step 1 Input sample dataset
 - $X = \{x_1, x_2, \dots, x_n\}$
 - $x_i \in R^m; i = 1, 2, \dots, n$
- step 2 Feature extraction based on KPCA
 - ① Calculate kernel matrix
 - $\tilde{k}_{ij} := (\phi(x_i), \phi(x_j)) = k(x_i, x_j); i, j = 1, 2, \dots, n$
 - $K = [k_{ij}]_{n \times n} = \frac{1}{n} \sum_{p,q=1}^n l_{ip} k_{pq} l_{qj} = \frac{1}{n^2} \sum_{p,q=1}^n l_{ip} k_{pq} l_{qj}$
 - ② Solve the eigenvalue problem
 - $[K - \lambda I] v = 0$
 - $(\lambda, v) = \frac{1}{\sqrt{\lambda_k}}; k = 1, 2, \dots, r$
 - $v_k = \sum_{i=1}^n \frac{\alpha_{ki}}{\sqrt{\lambda_k}} \Phi(x_i)$
 - ③ Data reconstruction
 - $t_k = (v_k, \Phi(x)) = \sum_{i=1}^n \frac{\tilde{\alpha}_{ki}}{\sqrt{\lambda_k}} (\tilde{\phi}(x_i), \tilde{\phi}(x))$
 - $= \sum_{i=1}^n \frac{\tilde{\alpha}_{ki}}{\sqrt{\lambda_k}} \tilde{\kappa}(x_i, x)$
 - $x_{rec} = \left[\sum_{i=1}^n \frac{\tilde{\alpha}_{1i}}{\sqrt{\lambda_1}} \tilde{\kappa}(x_i, x), \dots, \sum_{i=1}^n \frac{\tilde{\alpha}_{ri}}{\sqrt{\lambda_r}} \tilde{\kappa}(x_i, x) \right]^T$
- step 3 Separability promotion based on LDA
 - ① Reconstructed dataset
 - $X^{(i)} = [x_1^{(i)}, x_2^{(i)}, \dots, x_{N_i}^{(i)}]$
 - $x_j^{(i)} \in R^d; i = 1, 2, \dots, c, j = 1, 2, \dots, N_i$
 - ② Calculate the scatter matrix and solve the generalized eigenvalue problem
 - $S_w = \sum_{i=1}^c \sum_{j=1}^{N_i} (x_j^{(i)} - m_i)(x_j^{(i)} - m_i)^T$
 - $S_B = \sum_{i=1}^c N_i (m_i - m)(m_i - m)^T$
 - $S_B w = \lambda S_w w$
 - ③ Make a projection of $x_j^{(i)}$ on w
 - $y_j^{(i)} = w^T x_j^{(i)}$

3. AN IMPROVED DIRECTED ACYCLIC GRAPH SUPPORT VECTOR MACHINE

Support vector machine is derived from the work of Vapnik and his co-workers, and has a theoretical background in the VC dimension of statistical learning theory and executes structural risk minimization, thus providing a better generalization performance than the other machine learning methods, which always use the empirical risk minimization.

The SVM, originally designed for the binary classification problem, has a nice geometrical interpretation of discriminating one class from the other by a hyperplane with the maximum margin. If the data cannot be separated by a hyperplane in the original input space, SVM performs classification by mapping the input data from original space to a high dimensional feature space and then

constructing the hyperplane that optimally separates the data into two classes in the feature space.

3.1 Principle of binary support vector machine

Given a training set $T = \{(x_1, y_1), \dots, (x_l, y_l)\} \in (R^n \times Y)^l$, where l is the number of samples, $i = 1, 2, \dots, l$. We refer to $x_i \in R^n$ as the i th sample and $y_i \in Y = \{-1, 1\}$ as its label. The classification problem is to find the hyperplane in a high dimensional feature space H , which divides the sample set in H such that all the points with the same label are on the same side of the hyperplane. SVM is to construct a map $\phi(\cdot)$ from the input space R^n to a high dimensional feature space H and to find an optimal hyperplane (w, b) in H such that the separation margin between the positive and negative examples is maximized. Mathematically, the SVM classification amounts to finding a weight vector w and a threshold b satisfying

$$\min \frac{1}{2} w^T w + c \sum_{i=1}^l \xi_i \tag{24}$$

$$s.t. \begin{cases} y_i [\phi(x_i) + b] \geq 1 - \xi_i \\ \xi_i \geq 0 \end{cases}$$

where $i = 1, \dots, l$, $c > 0$ is a regularization parameter for the tradeoff between model complexity and training error, and ξ_i measures the absolute difference between $w\phi(x_i) + b$ and y_i . Figure 1 gives a binary classification SVM graphical illustration. The balls and squares stand for two separable samples, H is the optimal hyperplane, H_1 and H_2 are the convex hull of each class samples respectively.

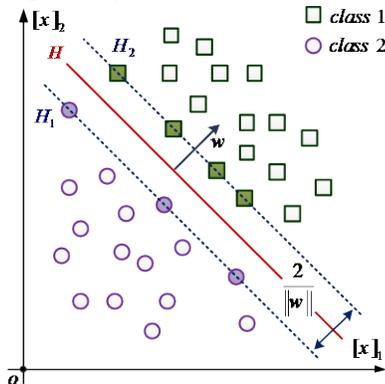


Figure 1: Principle of binary classification SVM

The Lagrange dual format of (24) is easier to solve

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j \kappa(x_i, x_j) \alpha_i \alpha_j - \sum_{i=1}^l \alpha_i \tag{25}$$

$$s.t. \begin{cases} \sum_{i=1}^l \alpha_i y_i = 0 \\ 0 \leq \alpha_i \leq c \end{cases}$$

where $i = 1, 2, \dots, l$, $\kappa(x_i, x_j)$ is the kernel function that satisfies $\kappa(x_i, x_j) = \phi(x_i)^T \phi(x_j)$.

Therefore, the learning problem in SVM is equivalent to the convex quadratic programming problem in (25). We have the decision function

$$f(x) = \text{sgn}[g(x)] \tag{26}$$

$$= \text{sgn}[\sum_{i=1}^l y_i \alpha_i \kappa(x_i, x) + b]$$

where $b = y_j - \sum_{i=1}^l y_i \alpha_i \kappa(x_i, x_j)$, $\alpha = (\alpha_1, \dots, \alpha_l)$.

3.2 Directed Acyclic Graph Support Vector Machine and its improvement strategies

SVM is originally designed for binary classification. However, the practical issues often require the discrimination for more than two categories. How to effectively extend it for multiclass classification is still an ongoing research issue.

There are two methods to construct multi-class SVM: direct methods and indirect methods. Direct methods include k -SVM and CS-SVM. Indirect methods include one-against-one support vector machine (OAOSVM), one-against-rest support vector machine (OARSVM) and directed acyclic graph support vector machine (DAGSVM). Direct methods are more intuitive, but the objective functions and the solution processes are complex, and have no advantages on classification accuracy, which make direct methods rarely used recently. Hsu [14] pointed out that DAGSVM was more suitable for practical use, compared with OARSVM and OAOSVM. Because DAGSVM has faster test speed than OARSVM and OAOSVM, and has no unclassifiable region. However, it still has the following shortcomings: Firstly, there are no guiding principles for the selection of the root node, which affect the classification accuracy. When different SVM is select as the root node, the classification results may be different, which makes the classification results uncertain; Secondly, when the DAGSVM is adopted to deal with Multi-classification problem, the error accumulation phenomenon caused by the inherent hierarchical structure will affect the classification accuracy. In

order to overcome the two shortcomings, a novel nodes arrangement method of DAGSVM based on fisher separability measure in feature space and margin of SVM is proposed in this paper.

DAGSVM was first proposed by Platt and Cristianini. It is a multi-class SVM based on OAOSVM, a brief introduction of OAOSVM is necessary. Its training phase is the same as the one-against-one method by solving binary SVMs. However, in the testing phase, it uses a rooted binary directed acyclic graph which has internal nodes and leaves. Each node is a binary SVM of i th and j th classes.

Given a test sample x , starting at the root node, the binary decision function is evaluated. Then it moves to either left or right depending on the output value. Therefore, we go through a path before reaching a leaf node which indicates the predicted class.

For an M -class problem, OAOSVM requires $M(M-1)/2$ bin-class SVM, and each of them is constructed by the training samples from two classes from M classes. The idea of OAOSVM is as follows.

Given a training set $T = \{(x_1, y_1), \dots, (x_l, y_l)\}$, where $x_i \in R^n$, $i = 1, 2, \dots, l$, $y_i \in \{1, 2, \dots, M\}$. All sample points $y = i$ and $y = j$ were extracted to constituted a new training set T_{i-j} at first. i and j satisfied $(i, j) \in \{(i, j) | i < j, j = 1, 2, \dots, M\}$.

Then the resolved decision function between class i and class j by SVM

$$f^{i-j}(x) = \text{sgn}(g^{i-j}(x)) \quad (27)$$

Consider classification results of all the $M(M-1)/2$ SVM about unknown sample x at last, the classification of sample x is decided by the votes of SVM on the result.

The topological structure diagram of DAGSVM for five classification issue is shown in Figure 2. From the topological structure diagram, we can see that DAGSVM is equivalent to operating on a list, where each node eliminates one class from the list. The list is initialized with a list of all classes. A test point is evaluated against the decision node that corresponds to the first and last elements of the list. If the node prefers one of the two classes, the other class is eliminated from the list, and the DAG proceeds to test the first and last elements of the new list. The DAGSVM terminates when only one class remains in the list.

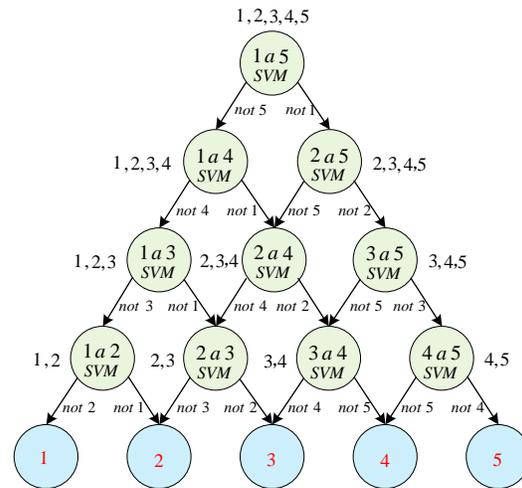


Figure 2: Topological Structure Diagram Of DAGSVM For Five Classes

A novel nodes arrangement method based on fisher separability measure in feature space and margin of SVM is proposed to improve the above-mentioned drawbacks of DAGSVM. In order to reduce the risk of misclassification, the class which is easiest to distinguish should be separated preferentially and the secondary easy-to-distinguish class should be separated closely followed, the classes which are difficult to distinguish should be separated at last. This operation makes sure that the classification error appears at the bottom layer as much as possible, and then leads to the reduction of decision risk and accumulative error. Node arrangement method based on inter-class Euclidean distance is adopted first. It is obviously that the larger of the Euclidean distance between two classes of samples, the higher of the classifier position. However, when there are cross-distribution phenomena, Euclidean distance cannot evaluate separability of the classes effectively. This situation is shown in Figure 3, Euclidean distances are the same in the two diagrams, but obviously the classes in the left diagram are easier to separate.

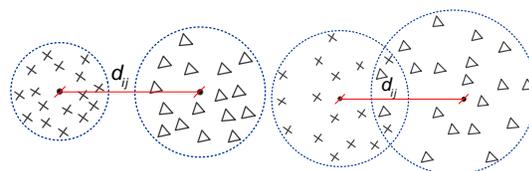


Figure 3: Comparison Of Between-Class Separability With The Same Euclidean Distance

Therefore, the fisher distance is a better choice. The definition of the fisher distance in Euclidean space is

$$d_{ij} = \frac{\|c_i - c_j\|}{\delta_i^2 + \delta_j^2} \quad (28)$$

where $\|c_i - c_j\|$ is the distance between the i th class and the j th class, and c_i is the class-center of the i th class, c_j is the class-center of the j th class.

SVM deals with data in high dimensional feature space actually. The form of distribution of the data will be changed when the data are mapped to high dimensional feature space F by a nonlinear map $\phi: x \mapsto \phi(x) \in F$ from the original input space X , where X is a subset of R^n . Therefore, fisher separability measure in high dimensional feature space must be given firstly.

Fisher separability measure in high dimensional feature space is defined as

$$D_{ij} = \frac{\|C_i - C_j\|}{\sqrt{\Delta_i^2 + \Delta_j^2}} \quad (29)$$

where C_i is the class-center of i th class in high dimensional feature space, C_j is class-center of j th class in high dimensional feature space, Δ_i^2 and Δ_j^2 represent variance high dimensional in feature space.

$$C_i = \frac{1}{n_i} \sum_{s=1}^{n_i} \phi(x_{is}) \quad (30)$$

$$C_j = \frac{1}{n_j} \sum_{t=1}^{n_j} \phi(x_{jt}) \quad (31)$$

$$\Delta_i^2 = \frac{1}{n_i - 1} \sum_{s=1}^{n_i} \|\phi(x_{is}) - C_i\|^2 \quad (32)$$

$$\Delta_j^2 = \frac{1}{n_j - 1} \sum_{t=1}^{n_j} \|\phi(x_{jt}) - C_j\|^2 \quad (33)$$

Because the explicit expression of ϕ is unknown, the formulas (30)-(33) cannot be solved directly. But, through the kernel function we can get D_{ij} directly without calculation of formulas (30)-(33).

Obviously, Fisher distance D_{ij} in feature space can be obtained by a known kernel function $\kappa(x, x')$, the separability of i th class and j th class can be judged. It can be seen from VC dimension and statistical learning theory that the larger the margin $\|w\|$ is, the better the generalization performance of SVM, meaning better separability of two classes. Phetkaew realized this point first, and used it on selection of classification node. Therefore, in order to further reduce the classification risk, margin of different classes

$2/\|w\|$ also should be considered.

For a k -class problem, take any combination of two in training phase like OAOSVM does, C_k^2 bin-class classifiers are constructed and each of them has a corresponding $\|w\|$. According to the formulas about separability measure in feature space, C_k^2 Fisher separability measure can be obtained. Consider margin $2/\|w\|$ and separability measure D , node arrangement principle of DAGSVM is given by formula (34).

$$M_i = \beta \frac{\sqrt{\sum_{i=1}^{C_k^2} \|w^i\|^2}}{\|w^i\|} + (1 - \beta) \frac{\|D^i\|}{\sqrt{\sum_{i=1}^{C_k^2} \|D^i\|}} \quad (34)$$

where $i = 1, 2, \dots, C_k^2$. $\beta \in [0.5, 1]$.

In order to reflect the importance of $\|w\|$, Figure 4 illustrates the reason for the setting range of β schematically. The separability measures are the same ($D = D'$), however $2/\|w\| > 2/\|w'\|$, thus the impact of margin is usually considered to be more important.

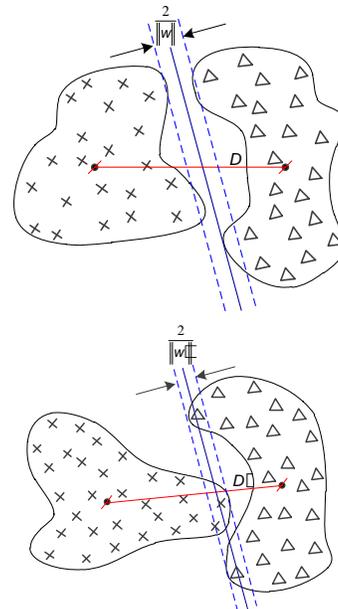


Figure 4: Comparison Of Margin With The Same Separability Measure

Firstly, M_i is arranged in descending order, then the node arrangement of DAGSVM can generally be determined by the sequence of M_i , from upwards to downwards and from left to right. The multi-class SVM constructed as the method described above is called improved DAGSVM

(IDAGSVM).

4. TWO-LEVEL STRUCTURE FOR FAULT DIAGNOSIS AND EXPERIMENTAL RESULTS

4.1 Two-level structure for fault diagnosis

The flowchart of analog circuits fault diagnosis based on the proposed method is given in Figure 5, which includes two procedures, namely, the training procedure and the diagnosis procedure. First, the Monte Carlo simulation is performed for the fault-free circuit and faulty circuit respectively and the output responses have been randomly divided into two parts including training samples and testing samples. Then, we extract the nonlinear fault features of the circuit under test (CUT) using separability promotion algorithm from the training set and construct the diagnostic model which is based on IDAGSVM.

When performing fault diagnosis, we extract the corresponding major fault features from testing set and import the features to the diagnostic model that has been constructed. Then the IDAGSVM can assort all the fault classes of the CUT.

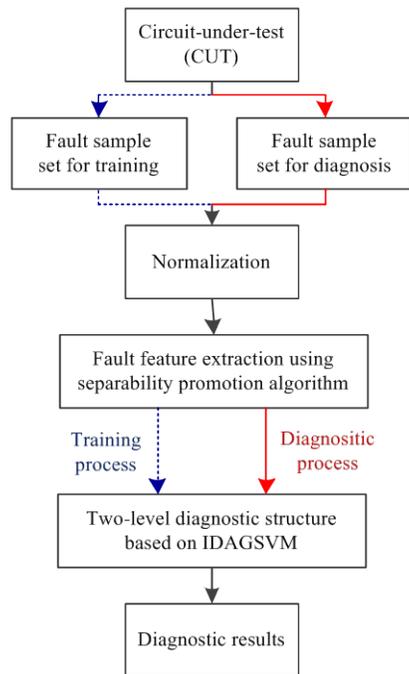


Figure 5: Comparison Of Margin With The Same Separability Measure

There are risks of false alarm and missing alarm in fault diagnosis system. We can tolerate the presence of the former to a certain extent, but usually we cannot accept the latter which could lead to tremendous loss. In order to reduce the risk

of false alarm and missing alarm, two-level diagnostic structure for analog circuit is proposed. In the first level, the training data is divided into two parts, including fault-free data (denoted as positive class points) and fault data (denoted as negative class points), then the labeled data is used to train the SVM. It is worth noting that such an approach will result in the unbalanced distribution of the training data in the first level. As shown in Figure 6, when the SVM is applied to dealing with class-unbalanced data, the generalized optimal separating hyperplane will be biased towards the majority class, thus causing poor diagnostic accuracy over the minority class.

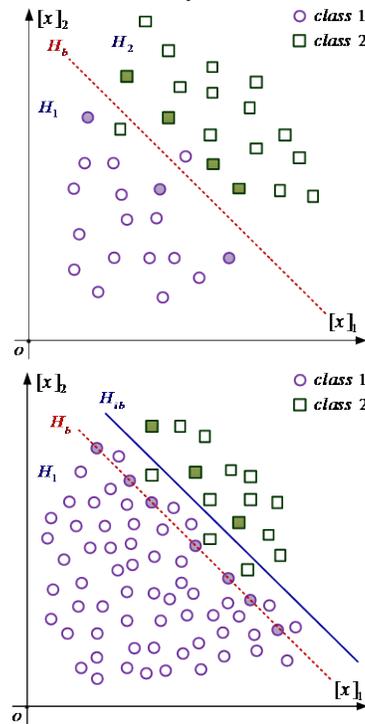


Figure 6: Unbalanced Classification Problem Based On SVM

With regard to analog circuit fault diagnosis, a certain degree of the movement for the optimal separating hyperplane is beneficial to diagnostic results. The reason is that we attach greater importance to false alarm as compared with missing alarm, while the excessive movement of the optimal separating hyperplane could have a detrimental effect on the ultimate diagnostic goal due to frequent phenomenon of false alarm. In order to deal with the classification problem of unbalanced data, the unbalanced SVM is adopted. The main idea of the unbalanced SVM is to introduce different loss functions for positively and negatively labeled points, which translates into a bias for larger multipliers for the class where the

cost of misclassification is heavier.

Mathematical description of the unbalanced SVM is as follows.

$$\begin{aligned} \min_{w,b,\xi} \quad & \frac{1}{2} \|w\|^2 + C_+ \sum_{y_i=1} \xi_i + C_- \sum_{y_i=-1} \xi_i \\ \text{s.t.} \quad & y_i((w \cdot x_i) + b) \geq 1 - \xi_i, \quad i = 1, 2, \dots, l \\ & \xi_i \geq 0, \quad i = 1, 2, \dots, l \end{aligned} \quad (35)$$

where C_+ is the error penalty for positive class, C_- is the error penalty for negative class.

The dual form of (35) is

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j \alpha_i \alpha_j \kappa(x_i, x_j) - \sum_{j=1}^l \alpha_j \\ \text{s.t.} \quad & \sum_{i=1}^l y_i \alpha_i = 0 \\ & 0 \leq \alpha_i \leq C_+, \quad \text{if } y_i = 1 \\ & 0 \leq \alpha_i \leq C_-, \quad \text{if } y_i = -1 \end{aligned} \quad (36)$$

By setting

$$\frac{C_-}{C_+} = \frac{l_+}{l_-} \quad (37)$$

the SVM that offers a better classification performance is derived, where l_+ is the positive class size, l_- is the negative class size. After the first level of the diagnostic structure is achieved, the second level can be determined by the sequence of M_i .

4.2 Diagnostic Flowchart

To verify the efficiency of the method proposed in this paper, ITC benchmark analog circuit leap-frog filter [15] is taken as the CUT. The circuit schematic and nominal values of the components are shown in Figure 7, the resistors and capacitors are assumed to have tolerances of 10% and 5%, respectively. Pspice tool is used to simulate this circuit.

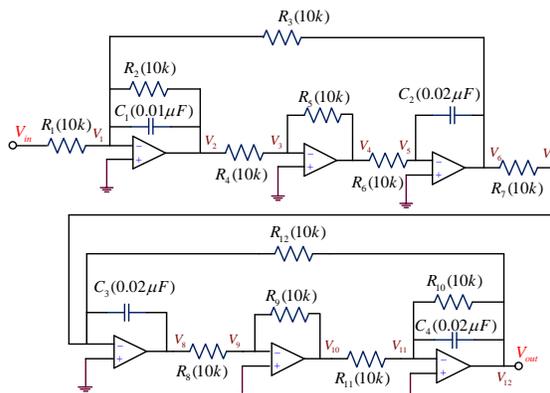


Figure 7: Leap-Frog Filter

All the parameters in the circuit are varied within the tolerance range and Monte Carlo simulation is performed for both fault-free and faulty circuit. After 200 Monte Carlo simulations, the output responses have been randomly divided into two parts including 120 training samples and 80 testing samples. The fault classes and the faulty component values are listed in Table 1, where \uparrow and \downarrow imply significantly higher or lower than nominal value.

Table 1: Fault Classes Of Leap-Frog Filter

Fault code	Fault class	Nominal value	Faulty value
F ₀	Fault-free	—	—
F ₁	C ₁ ↓	0.01μF	0.005μF
F ₂	C ₁ ↑	0.01μF	0.015μF
F ₃	C ₂ ↓	0.02μF	0.01μF
F ₄	C ₂ ↑	0.02μF	0.03μF
F ₅	C ₃ ↓	0.02μF	0.01μF
F ₆	C ₃ ↑	0.02μF	0.03μF
F ₇	C ₄ ↓	0.02μF	0.01μF
F ₈	C ₄ ↑	0.02μF	0.03μF
F ₉	R ₁ ↓	10kΩ	5kΩ
F ₁₀	R ₂ ↓	10kΩ	6kΩ
F ₁₁	R ₃ ↓	10kΩ	7kΩ
F ₁₂	R ₄ ↑	10kΩ	13kΩ
F ₁₃	R ₅ ↑	10kΩ	14kΩ
F ₁₄	R ₁₀ ↓	10kΩ	7kΩ
F ₁₅	R ₁₁ ↓	10kΩ	8kΩ
F ₁₆	R ₁₂ ↑	10kΩ	15kΩ

In our experiment, Gaussian kernel function $\kappa(x, y) = \exp(-\|x - y\|^2 / 2\sigma^2)$ is adopted and the selection range of regularization parameter c is $[2^{-4}, 2^{-2}, \dots, 2^{12}]$, the selection range of σ^2 in Gaussian kernel function is $[2^{-10}, 2^{-9}, \dots, 2^8]$. The sampled data are preprocessed by separability promotion algorithm for nonlinear feature extraction. Then the nonlinear features extracted from the test samples together with their labels are fed into the two-level diagnostic structure whose outputs estimate the probabilities that input features belong to different fault classes. The contribution rates of the first 10 KPCs for KPCA are listed in Table 2.

Table 2 : The Contribution Rates Of Kpcs

Kernel principal	Eigenvalue	Contribution rate
KPC ₁	9.97	35.02%
KPC ₂	7.55	26.52%
KPC ₃	4.75	16.68%
KPC ₄	2.86	10.05%
KPC ₅	1.68	5.90%
KPC ₆	0.89	3.13%
KPC ₇	0.43	1.51%
KPC ₈	0.21	0.74%
KPC ₉	0.09	0.32%
KPC ₁₀	0.04	0.14%

It can be seen from Figure 8 that the cumulative percent variance (CPV) of the first 5 KPCs is 94.17% in excess of 85%. At this point, it can be accounted that most of the nonlinear feature information has been extracted. The scatter diagrams of all kinds of fault features extracted by separability promotion algorithm are shown in Figure 9-12.

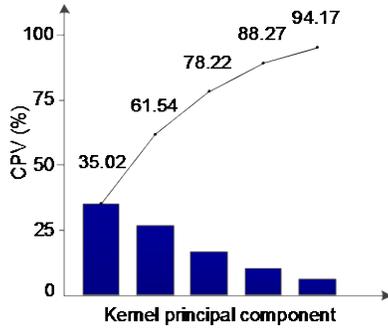


Figure 8: Cumulative Percent Variance

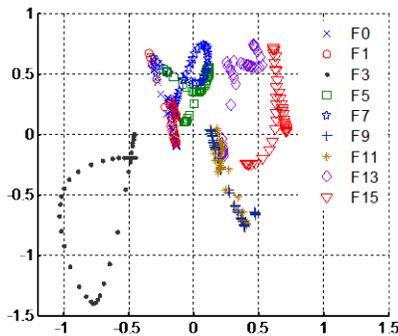


Figure 9: Two Dimensional Reconstructed Data Of Fault Classes With Odd Coding

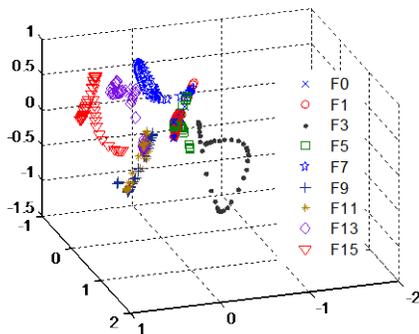


Figure 10: Three Dimensional Reconstructed Data Of Fault Classes With Odd Coding

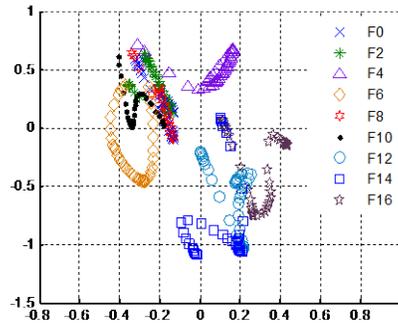


Figure 11: Two Dimensional Reconstructed Data Of Fault Classes With Even Coding

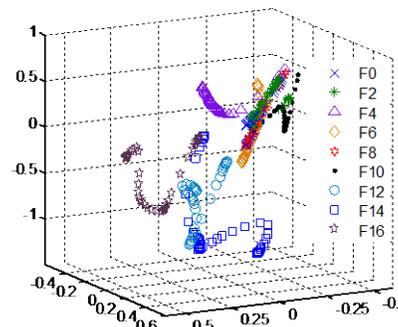


Figure 12: Three Dimensional Reconstructed Data Of Fault Classes With Even Coding

Table 3 : Diagnostic Accuracy Of The CUT

Fault code	OAOSVM	The proposed method
F ₀	72.50%	87.50%
F ₁	77.50%	92.50%
F ₂	82.50%	95.00%
F ₃	87.50%	100%
F ₄	76.25%	92.50%
F ₅	86.25%	100%
F ₆	83.75%	95.00%
F ₇	82.50%	100%
F ₈	73.75%	88.75%
F ₉	81.25%	97.50%
F ₁₀	88.75%	100%
F ₁₁	65.00%	96.25%
F ₁₂	82.50%	100%
F ₁₃	91.250%	100%
F ₁₄	78.75%	97.50%
F ₁₅	71.25%	95.00%
F ₁₆	86.25%	100%

For the convenience of comparison, the OAOSVM is adopted and the diagnostic results of the CUT are listed together in Table 3. It can be seen that the diagnostic accuracy obtained by the proposed method is higher than OAOSVM on any fault class. After simple calculation, the average diagnostic accuracy of OAOSVM without features extraction is 80.44%, while the average diagnostic accuracy of the proposed method is 96.32%.

5. CONCLUSION

Fault diagnosis of analog circuit is of essential importance for guaranteeing the reliability and maintainability of electronic systems. Two-level diagnostic structure for faulty analog circuit is proposed in this paper. In order to overcome the shortcomings of DAGSVM, the improved strategies based on fisher separability measure in feature space and margin of SVM is proposed. Then the separability promotion algorithm based on KPCA plus LDA is proposed and used to extract the nonlinear fault features. The experimental results show that it has a significant impact on analog circuits fault diagnosis, due to the excellent ability of nonlinear features extraction. Meanwhile, it can enhance class separability and reduce the within-class scatter in high dimensional space. Consequently, the purpose of improving the diagnostic accuracy is achieved.

ACKNOWLEDGEMENTS

This work was supported by NSFC (61074127).

REFERENCES:

- [1] H. Luo, Y. Wang, Jiang Cui. "A vague decision method for analog circuit fault diagnosis based on description sphere", *Chinese Journal of Aeronautics*, Vol. 24, No. 6, 2011, pp. 768-776.
- [2] F. Li, Y. P. Woo. "Fault detection for linear analog IC—The method of short-circuit admittance parameters", *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, Vol. 41, No. 9, 2002, pp. 105-108.
- [3] W. G. Fenton, T. M. McGinnity, L. P. Maguire. "Fault Diagnosis of Electronic Systems Using Intelligent Techniques: A Review", *IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Reviews*, Vol. 31, No. 3, 2001, pp. 269-281.
- [4] Y. Xiao, Y. He. "A novel approach for analog fault diagnosis based on neural networks and improved kernel PCA", *Neurocomputing*, Vol. 74, No. 7, 2011, pp. 1102-1115.
- [5] R. Spina, S. Upadhyaya. "Linear circuit fault diagnosis using neuromorphic analyzers", *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, Vol. 4, No. 3, 1997, pp. 188-196.
- [6] M. Aminian, F. Aminian. "Neural-network based analog circuit fault diagnosis using wavelet transform as preprocessor", *IEEE Transactions on Instrumentation and Measurement*, Vol. 51, No. 3, 2002, pp. 544-550.
- [7] M. Aminian, F. Aminian. "Neural network based analog circuit fault diagnosis using wavelet transform as preprocessor", *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, Vol. 47, No. 2, 2000, pp. 151-156.
- [8] Y. Xiao, L. Feng. "A novel neural-network approach of analog fault diagnosis based on kernel discriminant analysis and particle swarm optimization", *Applied Soft Computing*, Vol. 12, No. 2, 2012, pp. 904-920.
- [9] Y. Wang, J Cui. "A novel approach of analog circuit fault diagnosis using support vector machines classifier", *Measurement*, Vol. 44, No. 1, 2011, pp. 281-289.
- [10] H. Luo, Y. Wang, J. Cui. "A SVDD approach of fuzzy classification for analog circuit fault diagnosis with FWT as preprocessor", *Expert Systems with Applications*, Vol. 38, No. 8, 2011, pp. 10554-10561.
- [11] Z. Yang, Y. Peng, X. Peng. "Research on classification technique for imbalanced data set based on support vector machines", *Chinese Journal of Scientific Instrument*, Vol. 30, No. 5, 2008, pp. 1094-1099.
- [12] B. Schölkopf, A Smola. "Nonlinear component analysis as a kernel eigenvalue problem", *Neural Computation*, Vol. 10, No. 6, 1998, pp. 1299-1319.
- [13] H. Gao, J. W. Davis. "Why direct LDA is not equivalent to LDA", *Pattern Recognition*, Vol. 39, No. 5, 2006, pp. 1002-1006.
- [14] C. Hsu, C. Lin. "A comparison of methods for multiclass support vector machines", *IEEE Transactions on Neural Networks*, Vol. 13, No. 2, 2002, pp. 415-425.
- [15] J. Roh, J. A. Abraham. "Subband filtering for time and frequency analysis of mixed-signal circuit testing", *IEEE Transactions on Instrumentation and Measurement*, Vol. 53, No. 2, 2004, pp. 602-611.