The application of immune guided DTSVM in fault isolation

Yong Liua*, Aiqiang Xub, Yu Liu¹ a*

aGraduate Students’ Brigade, Naval Aeronautical and Astronautical University, Yantai 264001, China
b Department of Scientific Research, Naval Aeronautical and Astronautical University, Yantai 264001, China

Abstract

Support Vector Machine (SVM) is originally designed for linear binary pattern recognition. According to defect of common SVM, the Least square SVM is discussed to classify the small samples. In fault diagnosis, multi-category faults are common. Many multi-category SVMs are introduced in this paper briefly, and then an immune guided decision tree SVM (DTSVM) is presented which takes into account the fault occurring frequency. It proposes a fault isolation algorithm which adopts immune memory to optimize the configuration of DTSVM. Simulation and examples show that this method is useful in fault isolation. The kernel functions play an important role in the classification. The choice of kernel function and its parameters is neither easy nor trivial. In this paper, we have researched the kernel function of RBF and its effect on classification.

© 2011 Published by Elsevier Ltd. Open access under CC BY-NC-ND license.

Selection and/or peer-review under responsibility of [CEIS 2011]

Keywords: immune; DTSVM; fault isolation

1. Introduction

Fault diagnosis is still an ongoing research problem, for the equipments become more and more complex and the failure mode become more and more complicated, and the enormous improvements in the performance and cost of digital signal processing and communication devices in recent years[1]. The traditional approaches of fault diagnosis don’t present an easy solution. The artificial intelligence method can often obtain a better performance. Methods of expert system, artificial neural network, fuzzy
diagnosis, and so on, used commonly in the intelligent diagnosis at present, often need a large number of fault data samples or prior knowledge, but it is very difficult to obtain large number of typical fault data samples. Because a prior knowledge about the attribute is difficult to obtain, it is hard to select an appropriate fault diagnosis method. Support Vector Machine (SVM) is a new machine learning method developed in recent years, which requires relatively fewer learning samples. This property is very useful in fault diagnostics, which make it possible to compute directly using original data without preprocessing them to extract their features.

Artificial Immune System (AIS) is an emerging soft computing method inspired by natural immune system. The immune system is able to learn the structure of the pathogens, and remember those structures[2]. Future responses are much faster and, when made at an early stage of the infection, no adverse effects are felt by the organism. We underline the importance of this feature for smooth operation and cost savings, both in fault detection and in preventive maintenance. In this paper, an immune guiding multi-fault classifier based on SVM is proposed. The experiments’ results show that this method has good classification ability and high efficiency and is suitable for multi-fault diagnosis of equipments.

2. Support Vector machine

SVM is a relatively new soft computing method based on statistical learning theory presented by Vapnik (1995)[3]. In SVM, original input space is mapped into a high dimensional dot product space called feature space in which the optimal hyperplane is determined to maximize the generalization ability of the classifier. The optimal hyperplane is found by exploiting a branch of mathematics, called optimization theory, and respecting the insights provided by the Statistical Learning Theory[4].

Least squares SVM(LS-SVM) proposed by Suykens and Vandewalle (1999) are trained by solving a set of linear equations[5]. In contrast to the SVM, the LS-SVM is trained by minimizing

$$\min J(w, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{n} \xi_i^2$$

s.t. \(y_i(w \cdot \phi(x_i) + b) = 1 - \xi_i, \xi_i \geq 0, i = 1, 2, \ldots, n\)

In the LS-SVM, we use equality constraints instead of inequality ones employed in the conventional SVM. Therefore, the optimal solution can be obtained by solving a set of linear equations instead of solving a quadratic programming problem.

To derive the dual problem of (9), we introduce the Lagrange multipliers as follows:

$$L(w, \alpha, b, \xi) = \frac{1}{2} \|w\|^2 + \frac{C}{2} \sum_{i=1}^{n} \xi_i^2 - \sum_{i=1}^{n} \alpha_i \{y_i(w \phi(x_i) + b) - 1 + \xi_i\}$$

(10)

The conditions for optimality are derived by differentiating (11) with respect to \(w, \alpha, b, \xi\), and equating the resulting equations to zero:

\[w = \sum_{i=1}^{n} \alpha_i y_i \phi(x_i)\]

(11)

\[\sum_{i=1}^{n} \alpha_i y_i = 0\]

(12)

$$\alpha_i = C \xi_i$$

(13)

In a matrix form, (10), (11), (12) and (13) are expressed by
\[
\begin{bmatrix}
\Omega \\
Y \\
\end{bmatrix}
\begin{bmatrix}
\alpha \\
\beta \\
\end{bmatrix}
= \begin{bmatrix}
1 \\
0 \\
\end{bmatrix}
\]

(14)

Where \( \Omega, Y \) and \( 1 \) are, respectively

\[
\Omega_{i,j} = y_i y_j \phi^T(x_i) \phi(x_j) + \frac{\delta_{i,j}}{C}
\]

(15)

\[
Y = (y_1, y_2, \ldots, y_n)^T
\]

(16)

\[
1 = (1, 1, \ldots, 1)^T
\]

(17)

Here \( \delta_{ij} = \begin{cases} 
1 & i = j \\
0 & i \neq j
\end{cases} \).

Kernel function forms the lifeline of the SVM. The learning capability of SVM depends on the choice of kernel function. For a function to become kernel function, it has to satisfy Mercer’s Condition [6]. There are certain classes of kernel functions which satisfy Mercer condition. They are

1. Polynomial of degree \( d \)
   \[
   K(x, x_i) = [(x \cdot x_i) + \theta]^d
   \]

(18)

2. Radial basis function
   \[
   K(x, x_i) = \exp \left( \frac{-|x - x_i|^2}{\sigma^2} \right)
   \]

(19)

3. Sigmoid function
   \[
   K(x, x_i) = \tanh [u(x \cdot x_i) + r]
   \]

(20)

Each kernel function has parameters whose value has to be changed and tuned according to the data set.

3. Multi-class Classification Based on Immune

3.1. Multi-class SVM classification algorithm

SVM was originally designed for binary classification. Multi-class classification can be obtained by the combination of binary classification. There is a relationship between binary classification and multi-classification. Suppose we have a classable K-class problem; then it must be separated from each other by binary classification; On the contrary, in a K-class event, it must be K-class classable if any two classes of it is separable. We can construct a multi-class classifier by combining several binary classifiers. Several methods have been proposed[7]. Some of them are: “one-against-one”, “one-against-all” and “decision-tree support vector machine” (DTSVM).

For K-class event, the “one-against-one” method, construct \( M = C_i^2 = k(k-1)/2 \) classifiers, where each is trained on data from the \( i \)th and the \( j \)th class. Although the number is more than that of “one-against-all”, which needs \( k \) classifiers (as we know: \( k(k-1)/2 \geq k \) when \( k \geq 3 \), as each problem is smaller (only data from two classes). For the “one-against-all”, each classifier is trained with all the data, so the
total training time of “one-against-one” may not be more than that of “one-against-all” method (Hsu and Lin, 2002[8]). Decision tree support vector machine (DTSVM), which combines SVM and decision tree with the idea of dichotomy, is also an effective way for solving multi-class classification problems [9]. For a classification problem with m classes, DTSVM is a binary tree with k – 1 inner nodes that each node is a binary SVM classifier and k leaf nodes.

3.2. Immune guiding DTSVM algorithm

In DTSVM, the multiple classes are first classified into two sub-classes at the top node of decision tree, then the sub-classes will be classified once more, repeat this procedure until only a basic class is obtained, which is the leaf node of decision tree. Compared with “one-against-all” and “one-against-one”, DTSVM can resolve the unclassifiable regions, and has higher training speed since both the training sample number and class number are getting decreased from one node to the next. Therefore, DTSVM for multi-class classification has higher generalization ability than the two conventional methods. But a new problem exists in this method, i.e., the classification performance of DTSVM is highly related to its structure. Namely, one of the most important steps using DTSVM is to study how to determine a reasonable structure of decision tree, and so get the classification error minimized. To this end, according to the distribution of the multi-classes, based on the separability measured by Euclidean distance between the clustering centers of the two sub-classes, genetic algorithm is introduced into the formation of decision tree. The immune system is an efficient self-defense method that guards the human body from foreign antigens or pathogens[8]. AIS is an emerging soft computing method inspired by natural immune system. The literature[8] has proposed a new method for parameter optimization in SVM by a multi-objective artificial immune algorithm. In this paper, we try to use immune memory to optimize the hierarchy configuration of DTSVM. The purpose is to make the decision tree to obtain ability of machine learning.

The nodes division strategy of the decision tree is very important, which can improve the fault isolation effect. Here, we introduce the principle of immune response to save the fault occurrence frequency. Then fault samples are ordered by priority according to fault occurrence frequency. The top SVM is used to classify the fault pattern with highest occurring frequency. Next node is the SVM corresponding the lower priority sample. Like this, the whole nodes of the decision tree are decided by each sample. Although some faults have low occurring frequency, but it will be very dangerous if they occurred. Such faults could be treated in advance.

Figure 1 shows the topology of immune guiding DTSVM. In the figure, every sub-classifier is a LS-SVM using different parameters. The decision tree can be updated by the immune memory.

![Fig. 1. Topology of immune guiding DTSVM](image)
4. Experiments

4.1. Core function parameters selection

Kernel function and kernel parameters affect the performance of SVM. The selection quality of SVM parameters and kernel functions has an effect on the learning and generation performance. Appropriate kernel function and its parameters should be selected to obtain an optimal classification performance. In this paper, we only consider the RBF kernel using LS-SVMLab toolbox, the result is shown in figure 2.

\[
\text{LS-FSVM, } \gamma = 10, \sigma^2 = 0.01, \text{RBF, with 2 different classes}
\]

\[
\text{LS-FSVM, } \gamma = 10, \sigma^2 = 0.5, \text{RBF, with 2 different classes}
\]

\[
\text{LS-FSVM, } \gamma = 10, \sigma^2 = 100, \text{RBF, with 2 different classes}
\]

(a) \(\sigma^2 = 0.01\)  
(b) \(\sigma^2 = 0.5\)  
(c) \(\sigma^2 = 100\)

Fig. 2. Comparison of classification effect between different kernel parameters

The parameter \(\sigma^2\) reflects the characteristic distribution or range of training data. If the core parameter becomes smaller, the hyperplane is contracted around the learn sample points gradually. But if \(\sigma^2 \to 0\), the algorithm is equivalent to remember the isolated sample points. Therefore the system doesn’t have any ability of extension. On the other hand, if \(\sigma^2 \to \infty\), the algorithm will lose the ability of nonlinear processing.

4.2. Fault isolation

In this paper, we have compared “one-against-all”, “one-against-one” and “DTSVM” to isolate three kinds of fault. Kernel function is RBF and the parameter is chosen by the function “Tunelssvm()” automatically. In figure 3, (a), (b) and (c) represent the classification effect. Suppose the class 1, class 2 and class 3 are ordered by the fault occurrence frequency. The nodes of decision tree are configured by the priority of these three classes.

\[
\text{LS-FSVM, } \gamma = 1.3234, \sigma^2 = 1.3391, \text{RBF, with 3 different classes}
\]

\[
\text{LS-FSVM, } \gamma = 1.6254, \sigma^2 = 1.1232, \text{RBF, with 3 different classes}
\]

\[
\text{LS-FSVM, } \gamma = 25.9708, \sigma^2 = 2.5188, \text{RBF, with 3 different classes}
\]

(a) one-against-all  
(b) one-against-one  
(c) DTSVM

Fig. 3. Comparison of multi-class algorithm in fault isolation

The detail results of comparison of these algorithms are shown in table 1.

| Algorithm     | Classification Effect |
|---------------|-----------------------|
| one-against-all |                     |
| one-against-one |                     |
| DTSVM         |                     |

Table 1. Comparison of experiment data of 3 kind multi-class algorithms
From the table 1, we can see that DTSVM has less sub-classifiers. This algorithm uses immune memory to optimize the nodes of decision tree, so computing time is reduced than other algorithms.

5. Conclusion

The traditional intelligent fault diagnosis methods have worse training generalization especially in limited samples. The fine performance of SVM to limited samples attracts attention of investigations in fault isolation field. Due to the traditional multi-classification methods cannot solve the unclassifiable region and class clusters together, we proposed a new method named DTSVM which can improve the classification accurate rate. The immune memory is proposed to optimize the hierarchy configuration of DTSVM. Experiments results have proved its validity.

Acknowledgements

This work is supported by the Weapon Equipment Advanced Research Foundation of PLA (NO. 9140A25070208JB1402)

References

[1] M. Hajiaghajani, H. A. Toliyat, and I. M. S. Panahi, Advanced Fault Diagnosis of a DC Motor, IEEE Trans. Energy Conversion, 19 (1), 2004, pp. 60-65.
[2] D. Dasgupta, S. Yu, F. Nino, Recent Advances in Artificial Immune Systems: Models and Applications, Applied Soft Computing, 11(2), 2011, pp. 1574-1587.
[3] Fenton, W.G., McGinnity, T.M., Maguire, L.P., Fault diagnosis of electronic systems using intelligent techniques: a review, IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 31(3), 2002, pp. 269-281.
[4] N. Cristianini and J. Shawe-Taylor, An Introduction to Support Vector Machines and other kernel-based learning methods, Cambridge University Press, 2000.
[5] Suykens J.A.K, Vandewalle J, Least square support vector machine classifiers, Neural Processing Letters, 9(3), 1999, pp. 293-300.
[6] Ratnanjali Sood, The effect of kernel function on classification, In: Proceedings of National Systems Conference, 2008, pp. 369-373.
[7] Hsu, C.W., Lin, C.J., A comparison of methods for multiclass Support Vector Machines, IEEE Trans on Neural Networks, 13, 2002, pp. 415-425.
[8] K. P. Bennett, J. A. Blue, A support vector machine approach to decision tree, In: Proceedings of the IEEE World Congress on Computational Intelligence, Anchorage, AK, USA, 1998, pp.2396–2401.
[9] Ilhan Aydin, Mehmet Karakose, Erhan Akin, A multi-objective artificial immune algorithm for parameter optimization in support vector machine, Applied Soft Computing, 11(1), 2011, pp. 120-129.