

# Study on fault diagnosis method for nuclear power plant based on hadamard error-correcting output code

Y Mu\*, G M Sheng<sup>2</sup> and P N Sun<sup>1</sup>

<sup>1</sup>College of Physics Science and Technology Heilongjiang University, Harbin, China

<sup>2</sup>School of Electrical Engineering, Suihua University, Suihua, China

\*Corresponding author E-mail: muyufffff@126.com

**Abstract.** The technology of real-time fault diagnosis for nuclear power plants(NPP) has great significance to improve the safety and economy of reactor. The failure samples of nuclear power plants are difficult to obtain, and support vector machine is an effective algorithm for small sample problem. NPP is a very complex system, so in fact the type of NPP failure may occur very much. ECOC is constructed by the Hadamard error correction code, and the decoding method is Hamming distance method. The base models are established by lib-SVM algorithm. The result shows that this method can diagnose the faults of the NPP effectively.

## 1. Introduction

The technology of real-time fault diagnosis for NPP has great significance to improve the safety and economy of reactor. At present, many diagnostic methods are successfully used in nuclear power plant fault diagnosis [1-2]. As compared with the single classifier, the ensemble learning has higher accuracy and better generalization ability, so in recent years, the ensemble learning has been deeply researched in theory, and it has also been applied in the equipment fault diagnosis.

In the process of ensemble learning, how to generate a set of difference based models is a very important research content. And the methods, which are used to generate difference base models, include bagging, boosting, attribute combination and Error-correcting Output Coding (ECOC).

Different training subsets are obtained by extracting different subsets of attributes in attribute combination method. The author has done some work in this field. The model with good generalization ability is obtained by this method.

The model can still work properly in the absence of some parameters. However, this method is based on the fact that the parameters are redundant. But nuclear power plant is a complex system, Many kinds of faults may occur in the actual operation of nuclear power plants. So there are not so many redundant parameters. Therefore, in this paper, the sample set of the basic model by ECOC. The basic model is built by SVM algorithm.

## 2. SVM

SVM is an effective approach for pattern recognition. It is a machine learning method based on structural risk minimization principle[3]. SVM is a binary classifier, by solving a quadratic programming problem, we can find an optimal hyperplane that separates the two training data, and the SVM is constructed. The optimal hyperplane is that not only separates the two types of data correctly, but also makes the interval between two classes maximum. When a new sample  $x$  is introduced, its class is determined by the following decision function:



$$f(x) = \text{sgn} \left[ \sum_{i=1}^n y_i \partial_i K(x, x_i) + b \right] \tag{1}$$

Where  $\partial_i$  is Lagrange factor obtained by training, and  $x_i$  is support vector,  $y_i$  is the class label corresponding to  $x_i$ . And  $b$  is the classification threshold.  $K$  Is the kernel function. The kernel function is also the core of SVM. The linearly indivisible samples are mapped into high-dimensional space by the kernel function, which makes these samples linearly separable in high-dimensional space. SVM can effectively solve the problem of binary classification. At present, the commonly used multi-class SVM methods are one-versus-rest (1 V R), one-versus-one (1 V 1), Directed Acyclic Graph SVMs, Error Correcting Output Codes SVMs (ECOC –SVMs), etc. In this paper, ECOC algorithm is used to realize multi-class SVM.

### 3. ECOC

ECOC multi-class classification framework uses a binary or ternary coding matrix to achieve multi-class class decomposition and base classifier integration[4]. In the Coding matrix The binary code is represented by  $\{-1, +1\}$ , and ternary code is represented by  $\{-1, 0, +1\}$ . Where "-1" represents one class, "+1" represents another, and "0" indicates that the corresponding class of the codeword bit is ignored in the second-class partition formed. Figure 1 shows four common ECOC classification system diagram, namely: Figure 1(a) 1-V-R, Figure 1(b) 1-V-1, Figure 1(c) Dense Random, Figure 1(d) Sparse Random.

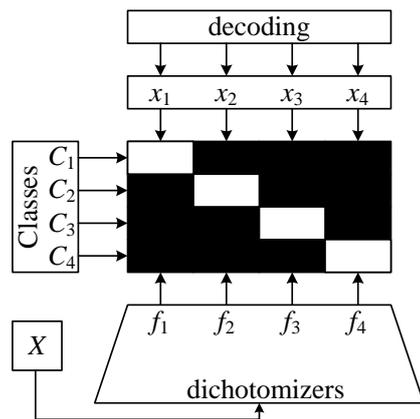


Figure 1( a) 1-V-R

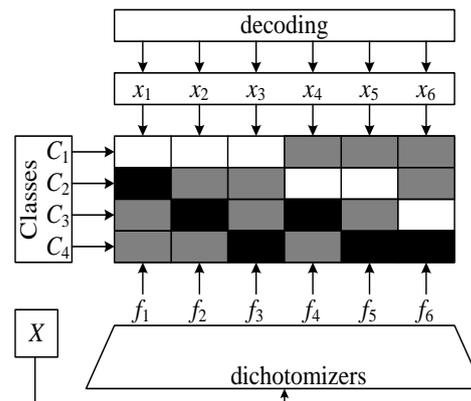


Figure 1( b) 1-V-1

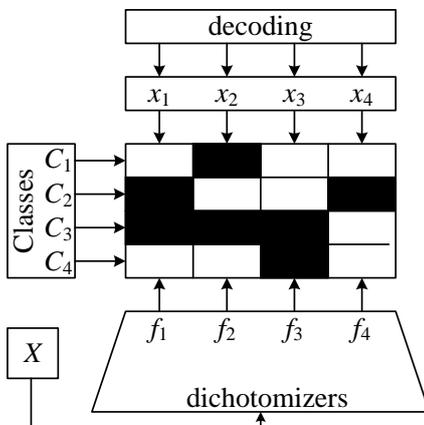


Figure 1( c) Dense Random

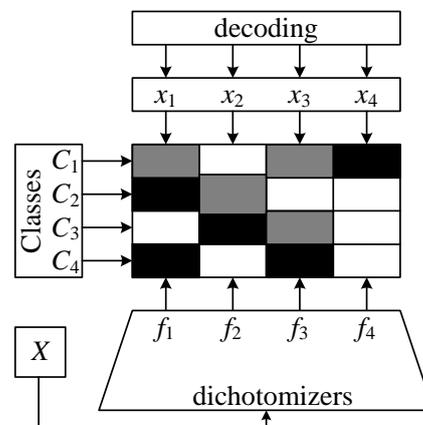


Figure 1( d) Sparse Random

Each row of the code array of Fig. 1 represents a codeword of a certain class, where  $C_i (i = 1, 2, 3, 4)$  is class label. Each column represents a second class partition of the sample. The symbols "1", "-1" and "0" are respectively expressed in white, black and gray. In the training phase, the training samples of each base classifier are transformed into two classes according to their corresponding columns in the coding matrix. And then two classes of classifiers corresponding to this column are trained respectively. In the test phase, a test sample  $X$  is given, and these classifiers are used to classify  $X$ . The result is a codeword vector  $(X_1, X_2, X_3, X_4)$ . Finally, the encoding matrix is decoded according to a certain decoding rule, and the final classification results are obtained by decoding. The key to multi-class classification using ECOC is to determine the effective coding matrix and decoding strategy.

### 3.1 Hadamard error correction code

The principle of how to design a good ECOC coding matrix is as follows:

1) The Hamming distances between each row are maximized as far as possible, so that the error correction ability of ECOC is stronger.

2) The Hamming distances between the column and the other columns and their supplements are maximized as much as possible, so that the difference between the base classifier is larger.

Based on the above principles, Dietterich and Bakiri give four kinds of ECOC coding methods: exhaustive codes, column selection from exhaustive codes, randomized hill climbing, BCH codes.

However, the error correction code obtained by this method has some drawbacks:

1) There is no algorithm for any class, and when the number of classes is changed, the algorithm must be replaced accordingly.

2) Some algorithms are very complex.

3) When the number of classes increased, the code length of the error correction code is longer. So It is difficult to store parameters and run in real time.

In this paper, ECOC is constructed by the Hadamard matrix. The Hadamard matrix is a binary code matrix. The order of the coding matrix can only be  $N = 2^i (i = 1, 2, \dots)$ . The second order Hadamard matrix is as follows:

$$H_2 = \begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix} \quad (2)$$

The higher-order Hadamard matrix can be recursively derived from the lower-order Hadamard matrix.

$$H_N = \begin{bmatrix} H_{N/2} & H_{N/2} \\ H_{N/2} & -H_{N/2} \end{bmatrix} \quad (3)$$

In the formula,  $-H_{N/2}$  means the complement of  $H_{N/2}$ . Complement operation, that is, elements "-1" and "1" interchange. The biggest characteristic of Hadamard matrix is that any two row or two columns are orthogonal. For the Hadamard matrix of order  $N$ , the Hamming distance between each row is  $N/2$ . So the Hadamard matrix has good discriminative performance. However, since its first row is all "-1", and the order of the coding matrix can only be  $N = 2^i (i = 1, 2, \dots)$ , so in practice, the Hadamard matrix needs to be modified in order to get a satisfactory Hadamard error correction code

Step 1 Determine the order of the Hadamard matrix. If  $2^{i-1} < k \leq 2^i$ , the matrix order is  $2^i$ .

Step 2 The Hadamard matrix is generated. The Hadamard matrix  $H_{2^i}$  of order  $2^i$  is generated according to (3).

Step 3 the first column of the Hadamard matrix  $H_{2^i}$  which all elements are are "-1" is deleted. A  $2^i \times 2^{i-1}$  matrix is obtained.

Step 4 The first  $k$  rows of the resulting matrix are taken according to the class number  $k$ . The error correction code matrix which is  $k \times 2^{i-1}$  matrix is obtained.

The error correction code matrix can be generated by the above method corresponding to the multi-class classification problem of any class. Simple and fast, and the code length is shorter. So the storage capacity is reduced and the operating efficiency is improved.

### 3.2 Decoding strategy

Hamming distance decoding is one of the most common decoding methods; its decoding rule is simple, high accuracy, and has a certain error correction capability. In this paper, Hamming distance method is selected as the decoding method.

## 4. Simulation Analysis

The typical faults of NPP which are chosen for analysis are as follows: loss of coolant accident (LOCA), feed water pipe rupture, steam generator tube rupture (SGTR).

The sample set is as table 1. The sample set includes training sample set and test sample set. There are 48 characteristic parameters of NPP, such as water level of pressurizer, flow, pressure and temperature of main steam etc.

**Table1.** Information of Sample Set

Class Label	Fault Type	Train Sample Number	Test Sample Number
1	Normal	80	80
2	LOCA (Cold Leg)	42	28
3	LOCA(Hot Leg)	43	28
4	SGTR	59	69
5	Main Steam Pipe Rupture	48	67
	total	272	272

The magnitude of the parameters varies widely. So the parameters must be normalized. At First the parameters are linearly normalized as follows:

$$x_{ij} = (x_{ij} - x_{\min j}) / (x_{\max j} - x_{\min j}) \quad (4)$$

The sample set contains data on 4 faults and normal operating states. The error correction code matrix of 5 classes is as (5)

$$\begin{matrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{matrix} \quad (5)$$

We test 1v1 method and ECOC algorithm, the diagnostic results is shown in Table 2.

The test results show that ECOC algorithm based on Hamming codes has higher accuracy, but the difference is not great. Since we only collected four classes of faults, so the sample set contains only five classes. In the 5-class problem, 1V1 algorithm needs to build 10 base models, and Hamming code ECOC algorithm to build 7 base models. Therefore, when the number of classes is small, there is not much difference between 1V1 algorithm and ECOC algorithm in the implementation of the efficiency. When the actual application of the fault diagnosis model is established, the number of classes

increased. The algorithm used in this paper will show the obvious advantages of the implementation of efficiency.

**Table2.** Fault Diagnosis Results

	1V1	ECOC
Classification accuracy (total)/%	90.44	91.18
Classification accuracy (class1) /%	100	100
Classification accuracy (class2) /%	89.29	89.29
Classification accuracy (class3) /%	71.43	75.00
Classification accuracy (class4) /%	84.29	85.51
Classification accuracy (class5) /%	94.03	94.03

## 5. Summary

The fault samples are collected from the NPP simulator of QINSHAN NPP Unit 1. ECOC is constructed by the Hadamard matrix, and Hamming distance method is selected as the decoding method. The base models are established by lib-SVM algorithm. The test results show that ECOC algorithm based on Hamming codes has a little higher accuracy. When the number of classes increased, the ECOC algorithm based on Hamming codes will show the obvious advantages of the implementation of efficiency. So the algorithm has a certain practical value, and it is very suitable for nuclear power plant fault diagnosis.

## Acknowledgement

This work was supported by the Natural Science Foundation of Heilongjiang University, China (QL201604).

## References

- [1] Moshkbar-Bakhshayesh K. Ghofrani M.B., Transient identification in nuclear power plants: A review, *Progress In Nuclear Energy*, 2013, 67: P23-32
- [2] Jianping Ma, Jin Jiang, Applications of fault detection and diagnosis methods in nuclear power plants: A review, *Progress in Nuclear Energy*, 2011, 53: P255-266
- [3] YIN Anrong, XIE Xiang, KUANG Jingming, Application of Hadamard ECOC in Multi-Class Problems Based on SVM. *Acta Electronica Sinica*. 2008,36(1): 122-126
- [4] Lei Lei, Wang Xiao-Dan, Luo X, Zhou Jin-Deng, Chen Qin, An overview of multi-classification based on error-correcting output codes, *Acta Electronica Sinica*. 2014,42(9): 1794-1800
- [5] T G Dietterich, G Bakiri. Solving multiclass learning problems via error-correcting output codes[ J ].*Journal of Artificial Intelligence Research* 1995, 2:263 -286.