wjert, 2017, Vol. 3, Issue 3, 94 -103. Original Article ISSN 2454-695X

World Journal of Engineering Research and Technology
WJERT
www.wjert.org

SJIF Impact Factor: 4.326

METHOD OF MECHANICAL FAULT RECOGNITION BASED ON HADAMARD ECC SUPPORT VECTOR MACHINE

Qingchao Bu*

Northeast Petroleum University Mechanical Science and Engineering lnstitute, Daqing, *Heilongjiang*, China, 163318.

Article Received on 09/03/2017 **Article Revised on 30/03/2017** **Article Accepted on 19/04/2017**

***Corresponding Author**
Qingchao Bu
Northeast Petroleum University Mechanical Science and Engineering lnstitute, Daqing, *Heilongjiang*, China, 163318.

ABSTRACT

Aiming at the ploblem that the traditional error correction code can not determine the code length and the optimal order, applied Hadamard matrix to the error correcting codes, a fault recognition method based on Hadamard error correction code support vector machine was proposed. Compared with the traditional error correction code, Hadamard error correcting code was more simple and the codebook was more advantageous. The method was verified by the sample of different classes of UCI database, and compared with the traditional "one to many" methods. Meanwhile, the method was applied to the fault identification of transmission mechanism of reciprocating compressor. The results show that Hadamard error correction code support vector machine has higher recognition accuracy in the case of a small number of classifiers and good stability for samples of different categories.

KEYWORDS: Pattern recognition; Hadamard matrix; error-correcting outputcode; support vector machine; multi-classification.

INTRODUCTION

Pattern recognition and pattern classification, is a kind of intelligent human computer application analysis field, with 1940s and 50s the rise of computer artificial intelligence, pattern recognition has been developed rapidly in early 1960s, and became a new subject of.[1] Multi class classification is an important field of pattern recognition, and it is also an

important way to analyze large sample data. At present, the commonly used multi classification methods have one to one (OAO),[2] one to many methods (OAA),[3] directed acyclic graph (DDAG),[4] two tree multi classification (BT), etc. It is worth noting that the error correcting codes proposed by Bakiri in 1995 and Diet-terich (Error Correcting OutputCode) method, namely the use of error correcting codes, one containing two elements of the N * L matrix (N is the number of categories L for two yards long), the classification problem is extended to multi classification problem, the method with the other several methods of comparison, showing the superiority of[6] strong. However, the traditional error correcting code is not given code length, sequence and codebook structure for different problems, there is not a unified standard error correcting code this code generation, for different number of categories, algorithm of error correcting codes are also different. At the same time, this method also has some shortcomings, such as weak generalization ability, slow training speed, and so on.

In order to solve these problems, this paper proposes a general error correcting Hadamard matrix and error correcting codes with the code matrix, the structure formation of simple and applies to any number of categories, combining with the structure of support vector machine, using support vector machine in the small sample, nonlinear and has good performance in high dimension vector space, will it is extended to multi classification problems. Finally, the method is applied to the fault identification of the transmission mechanism of reciprocating compressor, and compared with the traditional "one to many" classification method, and the feasibility of the method is verified by the data.

Error correcting code

Error correcting code (ECOC) is a distributed output code proposed by Bose and Ray in 1960. It is a method to transform the multi classification problem into two categories of Chaudhuri [7]. N samples, each equipped with a length of L was only 0 and 1 of the binary encoding sequence, consisting of a N L column codebook training, each column in the 0 corresponding data into a category, 1 corresponding data into another kind. Two yuan to construct a classifier, and so on, until the L two yuan to create a complete classifier. Test sample.

The input to the L classifier, a two element sequence A1, A2, al, And then calculate the sequence and test samples of all kinds of the Hamming distance between codewords (Hamming distance [8]), the minimum distance corresponding to the category is the category

of this new sample. Compared with other algorithm with error correcting code correction ability, namely when the minimum Hamming distance is D, the correction ability of at least /2 (D-1), even if there is a /2 error (D-1) classifier, the system can still give the correct results. Therefore, for some incomplete data, error correcting codes can still classify them correctly. A good error correction code should have the following points:

(1) In order to make the error correcting codes have strong error correction ability, should ensure that the relevant encoding matrix lines, namely line and line between the Hamming distance to reach the maximum.

(2) The column is not associated with other complementary or additional columns, which makes them the Hamming distance between the maximum, to prevent the same or similar classifier.

(3) The length of the coding is certain, the Hamming distance between codes is bigger, and the generalization ability is stronger. The Hamming distance between codes must be longer, and the extension ability is worse.

Based on the above three points, the K classification problems, encoding length of L must meet log2K <L = 2K-1 -1, which is produced in a two yuan training L classifier, when codebook columns in the same or similar, two yuan of classifier training will be the same or similar effect of promotion. The commonly used method of error correcting codes: (Exhaustive codes). The detailed code, suitable for the number of categories for the 3-7 sample; listed in the code. The detailed, suitable for the number of categories for 8-11 samples; random. (Randomized Hill Climbing) climbing algorithm and BCH code, suitable for the number of categories more than 11 of the samples. But until now there is no perfect theory to determine the optimal classification performance and determine the length of the code and the optimal order of the method. Such as the detailed code, the number of categories and relations for the L= 2k- 1- code length 1, namely 6 types of sample to construct error correcting code 31, if the number of samples increases, the code length will be in accordance with the exponential growth of two yuan, the number of classifiers need to build is relatively large, the effects of training time and generalization ability. At the same time, the error correcting code does not have a general algorithm for different types of numbers, that is, when the number of categories change, the algorithm should be the corresponding transformation, the implementation of more complicated. In view of the above problems, this paper uses Hadamard matrix and error correcting code to solve these problems effectively.

Hadamard error correcting code support vector machine

Hadamard error correcting code [9] is obtained by Hadamard matrix combined with traditional error correcting codes, the maximum separation degree of between the properties of Hadamard matrix can be realized according to the code of the row and column, meet an error correcting code is the ability requirements. Two order Hadamard error correcting code:

$$H_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \quad (1)$$

The order of the Hadamard matrix is 2 or a multiple of, where the order used is N= 2J (j= 1,2),... And the higher order Hadamard matrix can be obtained from the lower order Hadamard matrix by [10]: $H_N = \begin{bmatrix} H_{N/2} & H_{N/2} \\ H_{N/2} & -H_{N/2} \end{bmatrix}$ (2)

- HN/2 and HN/2 complement each other, that is, 1 changed to 0, and the number of "0" to "1", such as:

$$H_4 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

According to the properties of the Hadamard matrix, any two rows or two columns of the Hadamard matrix are orthogonal, and the distance between each row and each column is Hadamard for the N matrix of order N/2. Therefore, Hadamard error correcting codes have good inter class differentiation. In addition, the separation degree of row and column satisfies the requirement of error correcting codes in multi classification problems. The by using the Hadamard matrix and the error correcting code with error correction, eventually constructed a matrix of K * 2j- 1, this method solves the problems mentioned above, the error correcting encoding fast generation, provide the basis for further study of Hadamard error correction code multiple classification. However, this method only applies the Hadamard matrix to the error correcting code, and only 0 of the first column is deleted in the process of construction. The relationship between the number of samples and the code length obtained by the method of[9] is shown in table 1.

Table 1: The relationship between the number of Hadamard error correcting codes and code length.

Class number	length
3-4	3
5-8	7
9-16	15
17-32	31
33-64	63
65-128	127

Can be seen from table 1, when the class number is increased, the code length will increase, and the number of categories long is constant in a certain range of code, such as When the number of categories from 33 to 64, the code length is 63, in other words, the number of categories in this range has changed, the length of the code unchanged. By literature.[11] the number of samples less easy segmentation, so for different categories of the number of samples in the distribution will have different classification accuracy, i.e. the number of categories in the range of 33-64, the number of categories is 33, the highest classification accuracy, with the increase of the number of categories of the classification accuracy decreases, when the class number is 64 when reached the lowest; if the number of categories will continue to increase, the code length into another stage; similarly, the number of categories is 65, the highest classification accuracy, and then gradually decreased, when the class number is 128 to the lowest, so the classification accuracy of different for different number of categories using the same code length of the code. That is, the method can not guarantee the consistency of the sample separation accuracy. In this paper, we construct a codebook which produces different code lengths with different class numbers. Specific steps are as follows:

(1) According to equation (2) to generate the Hadamard matrix, if the number of categories N between 2i-1 and 2I, 2i-1<N = 2I, then generate the order number Hadamard matrix 2i;

(2) Delete the first column (the first column is all 0, no practical significance);

(3) On the matrix before the N row, the first N-1 column, generate the desired N * N-1 error correction matrix.

The error correcting code matrix using only type (2) can be obtained, thus generating simple; this method gives the number of different types of cutting method, when the class number is changed, only need to change the content that can be intercepted, when the class number is I, the code length is I-1, the number of categories and equivalence relation the code length, the

code length is short, improves the speed of operation; according to the properties of Hadamard matrix, when the error correcting code code length is N-1 (intercept before the code length is 2i-1), between the Hamming distance is 2i-1, and the column does not appear among the same and complementary, to ensure the separability of lines the row and column separation.

Support vector machine (Support Vector Machine, SVM Cortes and Vapnik1995 Corinna) is first proposed, it solve the small sample, nonlinear and high dimensional pattern recognition shows many unique advantages, and is widely used in function fitting learning problems in [12]. The kernel function is the core of SVM, the current commonly used kernel functions: polynomial kernel function (Polynomial), radial basis function (Radial Basis, Function, RBF), multi-layer perceptron (Multi-Layer Perceptron, MLP). Because of the superior performance of radial basis function, it is often used in practical applications. As a combination of the VC dimension theory and the structural risk minimization principle in the statistical learning theory, SVM has the following advantages:

(1) It is suitable for the finite sample case, the goal is to get the global optimal solution under the current information, rather than the optimal value when the number of samples tends to infinity;

(2) The algorithm is transformed into a quadratic optimization problem, and the global optimal solution is obtained instead of the local optimal solution obtained in the neural network;

(3) The kernel function is used to convert the inner product of the high-dimensional feature space to the kernel function in the low dimensional space, which avoids the curse of dimensionality.

(4) The support vector machine (SVM) is based on the principle of minimizing the empirical risk and confidence range, which can realize the structural risk minimization and has good generalization ability.

Hadamard error correcting code support vector machine (SVM) is a combination of the above Hadamard error correcting codes and SVM, which makes use of the advantage of SVM in dealing with small samples. When the class number is N, in the codebook, each column corresponds to a SVM training, each column represents a class of "0", "1" on behalf of another class, constructs a SVM model, and so on, until the N-1 classifier is created. During the test, the test samples are input into N-1 classifier, a length of two yuan N-1 sequence with

the sequence of various types, calculating the Hamming distance between codewords, on behalf of the class of minimum distance is the final result of discrimination.

The Hadamard code used in SVM, effectively solve the shortcomings of the traditional error correcting code error correction capability is weak, the code is not clear; at the same time, the method has the advantages of simple structure and effective guarantee between the code between the columns and the separability, indirectly ensured differences between the SVM, meet a structure the ideal error correcting code requirements.

Verification analysis

In practice, there are many kinds of samples, and the change of the number of samples will affect the recognition performance. Were selected in the UCI database, Segment Vowel, Image Segmentation test, the first of 3 sets of data are normalized, linear to [0, 1]; use PCA dimensionality reduction method to reduce the dimension of the sample to improve the computing speed. Radial basis function is used as kernel function and genetic algorithm is used to optimize the kernel parameter and penalty parameter C. In order to verify the classification performance of the method, three groups of data were trained before the testThe number of samples and the number of test samples are unified, and the data set is shown in table 2.

Table 2: UCI data sets and their attributes.

Data set	Training sample number	Number of test samples	Class number
Segment	100	100	7
Vowel	100	100	11
Image Segmentation	100	100	19

As shown in Table 3, we can see from the data in Table 3 that the Hadamard error correcting code method has a higher classification accuracy, and can keep a better classification stability when the number of samples is changed. However, when the number of classes increases, the recognition ability of "one to many" decreases rapidly, which is due to the decrease of the degree of discrimination between the models. The greater the number of categories, the more obvious the advantages of the Hadamard error correcting code method, such as when the sample number is 19, the recognition rate of the Hadamard error correcting code method is 95.96%, and the one to many method is only a gap of 47%. In fact, "one to many" method can be regarded as a special case of error correcting codes, the code for its unit array, the

Hamming distance is 2 lines, so they don't have the ability of error correction, recognition ability is weaker than Hadamard error correction method.

Table 3: Recognition accuracy of different categories.

Data set accuracy	OAA	Method in this paper
Segment	94.64%	96.21%
Vowel	79.53%	95.84%
Image Segmentation	48.31%	95.96%

Practical application

The data used in this experiment are from the Daqing oilfield Limited by Share Ltd gas branch of the 2D12 compressor. In the experiment, 10 sensitive test point is arranged on the connecting rod parts of the compressor driving mechanism, simulate the transmission mechanism of head bearing, bearing the bulk of a total of four large gap position of fault, and field test the compressor normal state data, using the method of fault data measured by classifying the obtained classification accuracy. The use of Beijing Oriental INV306-6660 multi-channel intelligent data acquisition instrument for data acquisition, sampling frequency 50KHz, sampling time 4S, sampling 100 times under each condition.

Based on the characteristics of multi fractal in dealing with complex system and describing the local scale of signal, it is often used in many fields such as data analysis and fault diagnosis [13]. Reciprocating compressor of complex structure, multiple excitations and fault information mutual interference, a single sensor is difficult to extract accurate feature, multi sensor generalized fractal dimension matrix and multifractal method can reflect the formation of system state information more comprehensive and more accurate. At the same time, in order to improve the computing speed, we need to reduce the dimension of data processing, PCA dimension reduction method is a commonly used method of data dimensionality reduction. Therefore, the multi fractal theory is used to form the generalized fractal dimension matrix to analyze the vibration signal of multi sensor, and the dimension reduction method is used to reduce the dimension of the PCA, and the feature vector is extracted.

SVM has been widely used in pattern recognition, regression estimation and so on because of its good performance in data analysis. Lib SVM is a SVM toolkit designed by Professor Lin Zhiren of National Taiwan University. It has a good performance in data classification, parameter optimization and model training. [14] Kernel function is the core part of SVM, and it is also the main factor to determine its performance. This paper uses the widely used RBF

kernel function, the specific form: K (x, Xi) = exp{- R, X - Xi, 2}, R as the kernel parameter, C error penalty parameter, using genetic algorithm to optimize the parameters and support vector machine.

The above five kinds of working conditions in the selection of the 100 groups of data, normalized and reduced dimension processing, in each set of conditions to select the 70 groups of data as training samples, select the data of the 30 groups as test samples. The Hadamard error correcting code support vector machine is established and compared with the "one to many" method. The results are shown in table 4. As can be seen from table 4, the Hadamard error correcting code SVM has higher recognition accuracy, and the recognition performance is better than the one to many method.

Table 4: Accuracy.

	OAA	**Method in this paper**
accuracy	94.92%	96.74%

CONCLUSION

According to the characteristics of Hadamard matrix, matrix and error correction encoding combined, using a Hadamard error code two classification problem is extended to multi classification problems in the way, and put forward a kind of code of the interception, interception after the code length and sample category number linear correlation, to avoid because of the number of categories have different classification accuracy of different.

Method to select different number of categories in the UCI database of the samples is verified and compared with the "one to many" method, the experimental results show that Hadamard code can successfully support vector machine is extended to multi classification problems, and the ability of classification is better than the "one to many" method, at the same time when the number of categories change can keep good stability classification, can effectively improve the performance of recognition system.

Taking the reciprocating compressor transmission mechanism as the research object, this paper analyzes the fault form and uses this method to identify the fault. The results show that the Hadamard code method has a good ability of fault recognition, better than the "one to many" method; and the method has the advantages of simple structure, easy to create a classification model, is of great significance for further study on the fault diagnosis.

REFERENCES

1. Liu Di, Li Yaofeng. Overview of pattern recognition [J]. Heilongjiang science and technology, 2012; 28.
2. Kreberl U-Pair Wise. Classification and support vector machines [C] Advances in Kernel Methods Supprot Vector Learning. Cambrige, 1999; 255-268.
3. Botton L，Cortes C，Denker J． Comparison of classifier methods:a case study in handw riting digit recognition[C]． Proc of the 12thIAPR Int Conf on Pattern Recongniton，Jerusalerm，1994: 77-82.
4. J C-Platt，Cristianini N，J Shawe-Taylor． Large margin DAGs formulticlass [C] Advances in Neural Information Processing Sys-tems, 2000; 547-553．
5. Gjorgji Madzarov，Dejan Gjorgjevikj，Iyan Chorbev． A multi-class SVM classifier utilizing binary decision tree[J]． Informati-ca，2009; 33(2): 233-241．
6. T G Dietterich,G Bakiri.Solving multiclass learning problemsvia error-correcting output codes[J]. Journal of Artificial Intel-ligence Research, 1995; 2: 263-286.
7. Bose RC,Ray-Chaudhuri, D K. On a class of error-correcting bi-nary group codes[J]. Information and Control, 1960; 3(1): 68- 79.
8. Zhang Huanjiong, Wang Guosheng, Zhong Yixin. Computer engineering and application of [J]. Text similarity computing based on Hamming distance of, 2001; 19.
9. Yin Anrong, Xie Xiang, velio tralli. Hadamard code using support vector machine [J]. Journal of electronics in multi classification problem in, 2008; 01.
10. Zhang Yanzhong, Shen Wei. Fast Fourier transform and transform [M]. Beijing: Aviation Industry Press, 1989.
11. Li Honglian, Jiao Ruili, Fan Jing. Accuracy analysis of multi class classification of support vector machines [J]. Journal of Beijing Insititute of Machinery Industry, 2008; 02.
12. Vapnik V. The nature of statistical learning theory[M]. New York: Springer, 1995.
13. Haiyang, Xu Minqiang, Study on fault feature extraction of reciprocating compressor based on Multifractal and singular value decomposition[J]. vibration and shock, 2013; 23.
14. Hsu C，Lin C J． A comparison of methods for multi-classsupport vector machines[J]. IEEE Transactions on Neural Networks, 2002; 13(2): 415-425.