Journal of Computers Vol. 28, No. 5, 2017, pp. 94-104
doi:10.3966/199115992017102805009

Fault Diagnosis of Single Yaw Damper Utilizing Hierarchical Multi-class Classifier

Daochao Tang[1], Na Qin[1*], Weidong Jin[1] and Peizhen Xu[1]

[1] School of Electrical Engineering, Southwest Jiaotong University,
Chengdu 610031, Sichuan, China
dctang1129@outlook.com, qinna@swjtu.edu.cn

Received 27 June 2016; Revised 7 February 2017; Accepted 8 February 2017

Abstract. In this paper, we propose a hierarchical multi-class classification approach which is optimized for single yaw damper fault problems. This novel approach associates with an original process of fault detection and localization which is arranged into support vector machine (SVM) with binary tree architecture. In fault detection, the developed method based on concatenated One-Against-One SVMs can significantly reduce the miss rate. Then, the hierarchical structure is built via iteratively partitioning the car bogie structure during fault location. This algorithm is very appealing as it takes advantage of the decision tree architecture and of SVM. Furthermore, the selection of error penalty factor C affects the precision of SVM due to its ability to avoid over fitting. In this paper, the Bayesian error estimator (BEE) which describes the error in a Bayesian framework is applied to obtain the optimal value of C . The effectiveness of this approach is illustrated experimentally on CRH3 EMU vehicle system.

Keywords: Bayesian error estimator, fault diagnosis, hierarchical classifier, yaw damper

1 Introduction

When High-speed Electric Multiple Units (EMU) is running in high speed, the working state of yaw damper is closely related with the vehicle stability and safety [1-2]. Furthermore, the yaw damper fault may cause the train derailment accident. CRH3 EMU bogie yaw damper adopts redundant design, namely each bogie with four yaw dampers. The present studies for yaw damper mainly focus on performance optimization. However, the fault analysis for it is less concerned [3-4]. For yaw dampers, the oil spill and joints aging are the most common of failures [2, 4]. In the actual operation of High-speed train, the fault of single yaw damper frequently occurred which will lead to uneven distribution of the car body's torque [2]. Thereby, fault diagnosis for yaw damper is an import process which can avoid serious damage and accident during practical operation. And a fault diagnosis of yaw damper should be researched. In a smart structure, the fault diagnosis is classically decomposed into four steps: detection, location, quantification and prognosis [5-7]. Therefore, a machine condition monitoring system can be effectively used as a part of decision support tool. In this paper, we address the detection and location of single yaw damper fault.

For multi-class classification problem, lots of classifiers are available, namely, artificial neural network (ANN), Bayes Net (BN), decision tree, support vector machines (SVM), etc. [5]. Moreover, SVM is a state-of-the-art machine learning algorithm due to its high accuracy and good generalization ability [6, 8]. Several methods have been proposed to solve multi-class problems with SVMs, such as One-Against-One (OAO), One-Against-All (OAA) and Direct Acyclic Graph (DAG) SVMs. Ensemble empirical mode decomposition (EEMD) and SVM were applied in feature classification of high-speed train bogie [3]. Statistical features and SVM were applied in fault diagnosis of automobile hydraulic brake system [8]. The power quality classifier based on wavelet trans-form and SVM was researched [9]. However, for all above method, too many binary SVM need to train and test. Recently, the hierarchical

* Corresponding Author

structure method has been researched [10-13]. It employs binary SVMs step by step according a hierarchical structure. The hierarchical classification performance can depend on the tree structure where the root node includes all classes, and each sub-node contains the partial classes' information and classifier. As we all know, the selection of error penalty factor C affects the precision of SVM. For each binary SVM model, the best value of penalty factor C is selected by cross-validation (CV) [14]. However, in this paper, it's selected by Bayesian error estimator (BEE) which describes the error in a Bayesian framework [14-16].

With the advantage of hierarchical multi-class classifier and considering the characteristics of vibration signal for yaw damper under fault condition, this paper proposes a fault diagnosis method based on the hierarchical structure which is arranged into SVM with binary tree architecture. There are two major scientific contribution of this paper. First, the developed method based on concatenated One-Against-One SVMs is put forward to judge whether the system is fault or not during fault detection. Second, the hierarchical multi-class classifier based on decision tree architecture and SVM is built via iteratively partitioning the car bogie structure.

To evaluate the effectiveness of this fault diagnosis approach applied in this special case, the manuscript is carried out as follows: Section 2 describes the related works including SVM, hierarchical structure, BEE and the key research problem. Section 3 introduces the vibration datasets and feature extraction methods. Section 4 are experiments and results analysis. Some useful conclusions and further work are drawn in the last sections.

2 Related work

2.1 Support Vector Machines

Classic support vector machine is a set of supervised learning methods which is an effective classification method. It is binary classification technique which constructs a hyper-plane or set of hyper-planes in a high or infinite dimensional space. This hyper-plane has the largest distance to the nearest training data points of any class (so-called functional margin) [8-9].

Give a set of training samples $(X_i, y_i), X_i \in \Re^n, y_i \in \{-1, 1\}, i = 1, \cdots, m$. SVM requires to solve the following constrained optimization problem (1) on a set of training samples.

$$\begin{aligned} &\max_{\alpha} \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i}^{m} \sum_{j}^{n} \alpha_i \alpha_j y_i y_j \kappa(X_i, X_j) \\ &subject\ to: \sum_{i=1}^{m} \alpha_i y_i = 0 \\ &\qquad 0 \le \alpha_i \le C, i = 1, \cdots, m \end{aligned} \tag{1}$$

where C is the penalty factor. In SVM, kernel function $\kappa(X_i, X_j)$ can be used to map original datasets to kernel space where it might be easier to separate. It has both linear and nonlinear version according to the type of kernel function. A SVM classifier will label a new unknown point X utilizing the following decision function (2).

$$f(x) = \sum_{i=1}^{m} y_i \alpha_i \kappa(X_i, X) + w_0 \tag{2}$$

where the α_i and w_0 are the optimal hyper-plane parameters. As stated before, the penalty factor C is to admit some misclassified points in the training set to prevent the over fitting problem. C can take values in the range $0 < C < \infty$. A small C means a very flexible classifier. On the opposite, a high value C means a strict classifier. If $C = \infty$, it becomes a hard classification. In this paper, factor C is selected by BEE.

2.2 Hierarchical Structure Using Binary Decision Tree

As described before, a new approach called hierarchical classification approach has been studied in recent years. According to literature [12-13], hierarchical classification approach differ in three principal criteria. The first criterion is the type of hierarchical structure: DAG or Binary Decision Tree (BDT). The second criterion is how deep the classification in the hierarchy is performed: mandatory leaf-node prediction, non-mandatory leaf-node prediction. The last criterion is how the hierarchical structure is explored: top-down, Big-Bang, or only predicts the leaf nodes.

In this paper, we proposed a hierarchical multi-class classifier using binary SVMs. This method is arranged into a BDT structure [10]. In the tree structure, the tree is created such that the classes at each parent node are divided into two clusters, one for each child node. This process continues until the leaf nodes contain only single class. The top-down approach is applied to explore tree structure. In order to build the tree, at each node the remaining classer are separated in two subgroups. A SVM classifier decides to which subgroup the new point belong to.

The hierarchical multi-class classier components are:

Node. A root node contains all k classes' information. Leaf nodes represent only one class, and it is located at the end level. Internal nodes are not only the sub-node of upper layer, but also the parent node of next layer.

Level. This BDT structure has layers. A layer is represented by a level $n (n = 0,1,\cdots,n = (k-1)/2)$. There is only a root node at the top-level $(n = 0)$. There are k leaf nodes at the bottom level $(n = (k-1)/2)$.

Arc. Tree building for learning and tree exploration for classification are performed in top-down direction. All nodes in BDT are connected by arcs.

2.3 Bayesian Minimum Mean-square Error Estimator

The optimal value of C is traditionally selected by cross-validation (CV). A more recent approach, the Bayesian minimum mean-square error estimator for classification error describes the error. In literature [14-15], the Bayesian error estimator (BEE) can be more accurate than CV. In this paper, we will apply the BEE method to select the parameter C of SVM model. The classification errors are estimated directly from the training data, and splitting operations are not required. The BEE will be briefly introduced in this section.

Consider a binary classification problem with labels $y \in \{0,1\}$, where from each class has n_y samples points, $X_i^y, i = 1,\cdots,n_y$. We assume a parametric model assigning class y an unknown parameter θ_y, which defines a class conditional distribution $f_{\theta_y}(X \mid y)$. Given a prior probability $\pi(\theta_y)$, the posterior distributions $\pi^*(\theta_y)$ for the parameters are given by (3)

$$\pi^*(\theta_y) \propto \pi(\theta_y)\prod_{i=1}^{n_y} f_{\theta_y}(X_i^y \mid y) \tag{3}$$

where the constant of proportionality is found by normalizing the integral of $\pi^*(\theta_y)$ to 1. When the prior probability is proper, this follows from Bayes' rule; If $\pi(\theta_y)$ is improper, this is taken a definition. Let c (differ from the penalty factor C) be the a priori probability for class 0, and let ε_n^y be the error contributed by class y on some classifier [16]. Assuming c, θ_0 and θ_1 are independent, the BEE is given by (4)

$$\hat{\varepsilon} = E_{\pi^*}[c]E_{\pi^*}[\varepsilon_n^0] + (1 - E_{\pi^*}[c])E_{\pi^*}[\varepsilon_n^1] \tag{4}$$

where

$$E_{\pi^*}[\varepsilon_n^y] = \int \varepsilon_n^y(\theta_y)\pi^*(\theta_y)d\theta_y \tag{5}$$

and $E_{\pi^*}[c]$ is found according to our prior model for c when the prior probabilities are improper, this is called the generalized BEE.

In order to evaluate the integral (5), we assume Gaussian class-conditional densities $f_{\theta_y}(X \mid y)$ for the data. The parameters of the Gaussian model describe the feature label distribution and denoted by $\theta_y = \{\mu_y, \Sigma_y\}$, where μ_y is the mean with a parameter space, and Σ_y is the covariance matrix. Considering one class at a time, we assume Σ_y is nonsingular. In literature [16], the priors of the form (6):

$$\pi(\theta_y) \propto \left|\Sigma_y\right|^{-(\kappa+D+1)/2} \exp(-\frac{1}{2} trace(S \Sigma_y^{-1})) \left|\Sigma_y\right|^{-1/2} \times \exp(-\frac{\nu}{2}(\mu_y - m)^T \Sigma_y^{-1} (\mu_y - m)) \tag{6}$$

where κ is a real number, S is a non-negative definite $D \times D$ matrix, $\nu \geq 0$ is a real number, and m is a length D real vector.

In this paper, we assume that Σ_y^{-1} are general covariance matrices. In this case, $\pi^*(\Sigma_y)$ is an inverse-Wishart distribution (7):

$$\pi^*(\Sigma_y) = \frac{\left|S^*\right|^{\kappa^*/2} \left|\Sigma_y\right|^{-(\kappa^*+D+1)/2}}{2^{\kappa^* D/2} \Gamma_D(\kappa^*/2)} \exp(-\frac{1}{2} trace(S^* \Sigma_y^{-1})) \tag{7}$$

where Γ_D is the multivariate game function. The BEE can be obtained as formula (8):

$$E_{\pi^*}[\varepsilon_n^y] = \frac{1}{2} + \frac{\text{sgn}(A)}{2} I(\frac{A^2}{A^2 + \alpha^T S^* \alpha}; \frac{1}{2}, \frac{\kappa^* - D + 1}{2}) \tag{8}$$

where

$$A = (-1)^y g((m)^*) \sqrt{\frac{\nu^*}{\nu^* + 1}} \tag{9}$$

where $I(x; a, b)$ is the regularized incomplete beta function (10):

$$I(x; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^x t^{a-1} (1-t)^{b-1} dt \tag{10}$$

2.4 Classification Using Hierarchical Multi-class Classifier

In this section, the key research problem of this paper will been described. In section 2.2, the hierarchical structure has been described in detail. The hierarchical structure can be built by many different criteria which is one of key study. Genetic algorithm [11, 13] and Euclidean distance [12] have been applied to build structure. Considering the bogie has strict geometric structure, therefore, the hierarchical structure will be built by partitioning the car bogie structure. Let us consider a whole vehicle model which is divided into 8 sub-parts according to 3 forms of axes in Fig. 1. In this special case, 9 classes are considered: the 8 different positions single yaw damper fault (class 1-8) and the normal working condition (class 0). The first step of the tree is the detection step, where the classifier is to judge whether the system is fault. Then decisions are taken recursively to localize it, if a fault is detected.

In this part, we will pay attention to multi-class classification problems, i.e., classification using hierarchical structure. In the Fig. 1, the whole vehicle model is divided into 8 zones which represent 8 working conditions by 3 forms of axes. So, the hierarchical structure of BDT is illustrated in Fig. 2. The structure of BDT is mainly composed of two parts: fault detection and fault location which will be introduced in next section.

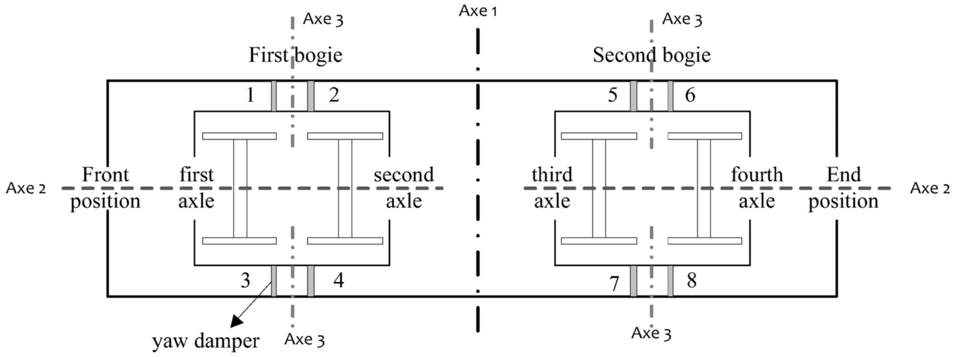

Fig. 1. The position of yaw damper in the bogie

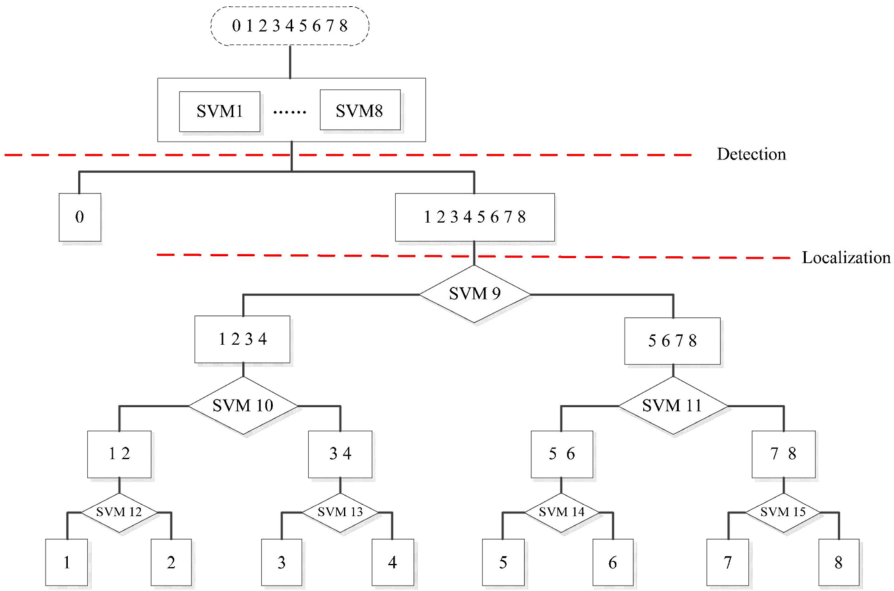

Fig. 2. Hierarchical structure

Fault detection. In practical application, we should first carry out fault detection to judge whether the system is fault or not. In the first step, we present a novel approach, aiming to identify the normal condition. Fig. 3 shows the developed classifier. The eight OAO binary SVM models are concatenated. And each one is that normal compared with one kind of fault working condition. Each node of this novel approach is parallel to each other. And they are also independent. This structure mainly includes two steps:

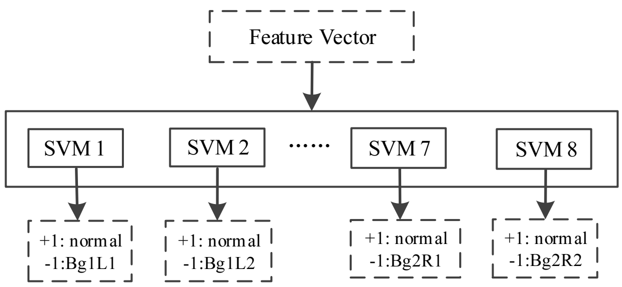

Fig. 3. Fault detection with OAO binary SVM

(1) Step 1: Training step, each OAO SVM model will be trained by normal and each fault condition's data in turn.

(2) Step 2: Decision step, a new unknown vector X will be predicted 8 times by each SVM model. Then all predicted labels will be summed. At last, if $sum = 8$, the new vector is normal or fault (we don't

need to know what kind of fault).

Multiple OAO SVM combination method can effectively avoid the phenomenon of deviation of hyper-plane. The purpose is to reduce the missing sample in the first step. It is difficult to find the optimal hyper-plane when all fault conditions are taken as negative samples and the normal as positive samples. On the contrary, it is easier that 8 decision functions will be achieved than the optimal one.

Fault location. The fault location is a BDT structure which takes advantage of both the tree architecture and SVM. Utilizing this architecture, $(N-1)$ SVMs needed to be trained for a N class problem, but only $\log_2 N$ SVMs are required to be consulted to decide a unknown sample. At each node of tree, a decision is made about the multi-class information by transferring the sample to the left or to the right sub-node.

According to the whole vehicle model shown in Fig.1, the fault location experiments include three steps. The first step is fault location between first and second bogie. All fault working conditions of first bogie will be as positive samples (class 1-4). On the contrary, the second bogie's conditions as negative samples (class 5-8). In next, fault location between right and left are conducted on the first bogie (class 1, 2 vs. class 3, 4) or the second one (class 5, 6 vs. class 7, 8). Finally, we will location which yaw damper has fault, .i.e., leaf nodes represent only one class which represents the 8 different positions single yaw damper fault.

3 Data Source and Features Extraction

3.1 Data Source

The simulation data is achieved using the nonlinear dynamic model of EMU vehicle system which is established with multi-body dynamics analysis software (SIMPACK). LMA wheel tread and CN60 rail are applied in dynamic model, and inner moment of wheel chooses China standard. Nonlinearity of wheel/rail contact, wheel/rail creep and suspension are taken into consideration. The vehicle model is constituted with one carbody, two bogie frames, four wheelsets, eight tumblers and two traction rods. The track excitation spectrum which was measured in Wuhan-Guangzhou passenger line is adopted during simulation experiment.

From the Fig. 1, there are eight yaw dampers of the whole vehicle. In this paper, 9 working conditions are researched for eight yaw dampers including normal, first left yaw damper of first bogie fault, second left one of first bogie fault, etc. The running speed of vehicle is 200 $km \cdot h^{-1}$ under each working condition. The vehicle is run for almost 10.0 min, and the frequency of sampling is 243 Hz. The vibration data of 58 channels can be gained from the whole vehicle model.

3.2 Features Extraction

Both ends of yaw dampers are respectively installed on the vehicle body and bogie frame. There are 18 channels in total on the vehicle body and bogie frame. Therefore, these 18 channels data are employed during next experiments. The sensors data mainly include vibration displacement and acceleration. The original signal is processed with following settings: (1) Sample length: To strike a balance, the number of samples is taken as 1024. (2) Number of samples: Since the vehicle is run for almost 10.0 min, 142 samples are taken for each working condition. Owing to limitations of space, Fig. 4 shows the part samples' time domain signal got from section 3.1.

In order to obtain the potential useful information, the 18 channels data are analyzed in the time domain, frequency domain and time-frequency domain. In the sequel, features are made up of three parts: dynamic performance indices, statistical parameters, and wavelet energy moments.

Dynamic performance indices. Some typical dynamic performance indices are selected as the part of feature vector. They are mean square deviation and Sperling indices. Passenger car's Sperling indices can be obtained by formula (11).

$$W = 0.986[\frac{A^3}{f}F(f)]^{0.1} \tag{11}$$

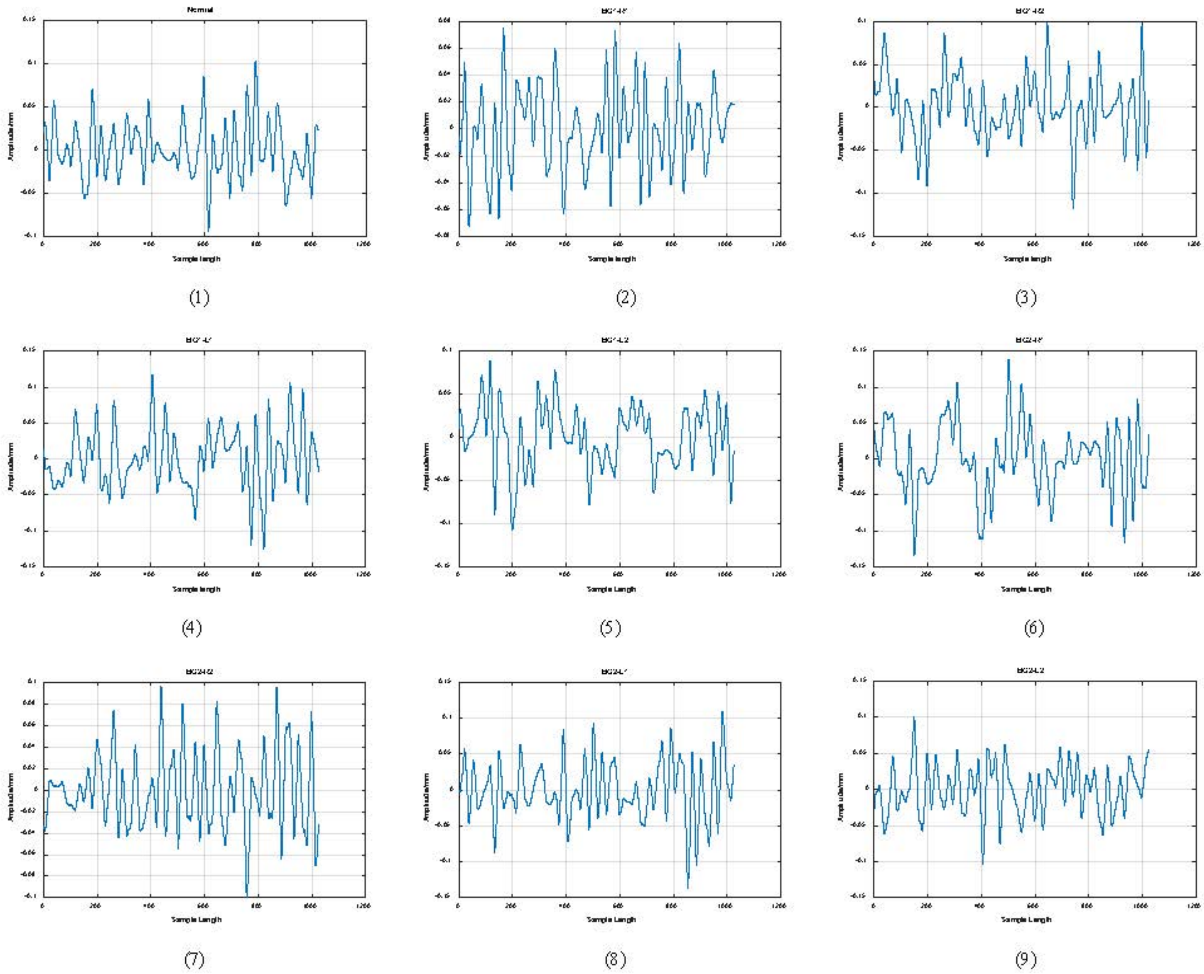

Fig. 4. Vibration signal of a single sample of various working conditions

where f is vibration frequency, A is vibration acceleration, and F is correction factor. Sperling indices of first two channels are computed in formula (13).

Statistical parameters. The statistical parameters are widely adopted for pattern recognition. Mean value, standard deviation, peak-peak value, peak value, mean square amplitude, root mean square, average amplitude, skewness factor, kurtosis factor, shape factor, peak factor, impuls factor, clearance factor, gravity frequency and frequency variance are extracted from the vibration signals.

Wavelet energy moments. In literature [3], the 3 layer wavelet packet transform and reconstruction were performed on the original signal $x(t)$. Then, the energy moment of third layer mth frequency band signal $S_{3m}(m=1,\cdots 8)$ can be computed through Eq. (12).

$$E_{3m} = \sum_{k=1}^{N} (k\Delta t) \left| S_{3m} k \Delta t \right|^2 \quad \textbf{(12)}$$

where Δt is sampling time interval, N is total sampling points, and m is order number of frequency band.

The whole vehicle is a unique system, we should analysis as a whole no matter what working condition. Given that each channel contains different parts of information, the signals are concatenated come from each channel. In this way, a 434-dimensional feature vector is obtained.

$$X_i = [f_1(c_1), f_1(c_2), f_1(c_1), \cdots f_{24}(c_1), \cdots, f_1(c_{18}), \cdots f_{24}(c_{18})] \quad i = \{1, \cdots m\} \quad \textbf{(13)}$$

where first two features are Sperling indices.

4 Experiments and Results Analysis

The fault diagnosis of the single yaw damper fault is taken up. The hierarchical approach is used with the features vector and C-SVMs. The 5-fold CV is adopted during the experiments and the number of training samples and test samples is listed in Table 1. In our work, we use the *Python* and *sklearn* to implement the hierarchical structure and C-SVM.

The C-SVM is trained using linear, RBF, Sigmoid and polynomial kernel functions. Kernels are compared with the accuracy while they are used for multi-class classification. For different kernel functions, the average classification accuracy of C-SVM is shown in Table 1. The penalty factor C is the default value (0.001).

Table 1. Average classification accuracy of hierarchical approach

	Type of kernel function	Training samples	Test samples	Average accuracy (%)
1	Linear	1026	252	**91.18**
2	RBF	1026	252	72.22
3	Sigmoid	1026	252	34.87
4	Polynomial degree 1	1026	252	65.08
5	Polynomial degree 2	1026	252	55.95
6	Polynomial degree 3	1026	252	41.80

Among all other kernel functions, the linear kernel for C-SVM obtains the best classification accuracy (91.18%). We can get the same conclusion as literature [16]. Due to a small number of training samples (compared to the data dimensionality), linear classifiers may outperform in such cases. Owing to that the highest average accuracy (91.18%) is gained with the linear C-SVM, the linear C-SVM model is mainly adopted in the following section.

4.1 Fault Detection Experiment

From section 2.4, fault detection is binary classification problem. Normal working condition's samples are positive sample, and fault samples as negative. In this section, some typical algorithms are chosen to compare with the proposed algorithm. The linear kernel function is selected during experiments and the penalty factor C is the default value. The experimental results are the average value of repeat 20 times. The 5-fold CV is adopted during the experiments

From Table 2, the other algorithms have emerged the phenomenon of deviation of hyper-plane. On the contrary, proposed algorithm in this paper have get much better results. Two fault sample is incorrectly identified as normal. The result in Table 2 proves that the miss rate of the hierarchical structure is fairly low.

Table 2. The result of fault detection

	Type of classifier	True positive rate (%)	True negative rate (%)	Accuracy (%)
1	KNN	53.57	89.73	85.71
2	BP neural network	50.00	93.30	88.49
3	Binary SVM	53.57	98.21	93.25
4	Proposed algorithm	**92.86**	**98.66**	**98.01**

4.2 Fault Location Experiment

The 8 different positions single yaw damper fault means that 8 different kinds of samples need to be processed. In other word, fault location experiment is a classical multi-class classification problem. In this section, some methods that have been proposed for solving multiclass problems with SVMs compared with hierarchical multi-class classifier which been put forward in this paper. The results that are the average value of repeat 20 times is shown in Table 3.

Table 3. The result of fault location

	Type of classifier	Train samples	Test samples	Accuracy (%)
1	OAO SVM	912	224	89.74
2	OAA SVM	912	224	87.49
3	Proposed algorithm	912	224	**91.51**

From Table 3, the hierarchical multi-class classifier achieves the better results than other classical two multi-class classification. However, the results that get by proposed algorithm are not the ideal. May be due to the parameters of the C-SVM model are chosen wrong.

4.3 Penalty Factor Selection Using BEE

As all known, the selection of penalty factor C affect the precision of C-SVM. In section 4.1 and 4.2, the value of C is only default. Therefore, some results are not ideal. Next, we will focus on the selection of penalty factor C with BEE approach which is recently introduced in section 2.3. According to the prior knowledge, C can take values in the range $10^{-5} \le C \le 10^{5}$. The change step of C is 0.001.

Due to the BDT structure is composed with 15 independent C-SVM, the specific value of C vary from one model to another. So, the selection of penalty factor will repeat 15 times for each C-SVM. For reasons of space, two typical error curves for different factor C are shown in Fig. 5.

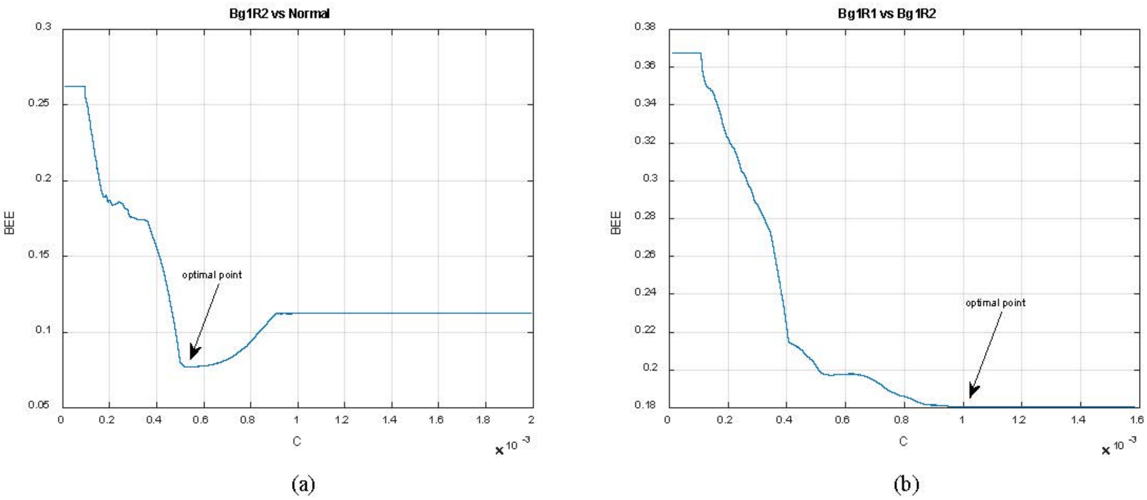

Fig. 5. Error curves for different C : (a) SVM 3; (b) SVM 10

Each C-SVM has its own optimal value which is listed in Table 4. From the table, the value of C is relatively small which means very flexible classifiers. After the optimal penalty factors are obtained, the next job is to retrain the hierarchical structure which composes with linear C-SVMs. And the 5-fold CV is also repeated for 20 times. All the other steps are the same as in section 4.1 and 4.2. The average classification accuracies for different optimal values' C-SVM are shown in Table 5. The confusion matrix for optimal values' C-SVM is also presented in Table 6. The hierarchical multi-class classifier of optimal values improves by 5% compared with default values. From the Table 6, we can also find that only 8 samples are misclassified. And there are no fault sample incorrectly identified as normal.

Table 4. Optimal penalty factors for different models

	SVM 1	SVM 2	SVM 3	SVM 4	SVM 5	SVM 6	SVM 7	SVM 8
Optimal C	0.0012	0.0012	0.0012	0.0005	0.0009	0.0006	0.0012	0.0008
	SVM 9	SVM 10	SVM 11	SVM 12	SVM 13	SVM 14	SVM 15	
Optimal C	0.0014	0.001	0.001	0.0006	0.001	0.0007	0.0015	

Table 5. Average classification accuracy for different optimal values' C-SVM

	Training samples	Test samples	Average accuracy (%)
Default values	1024	252	91.18
Optimal values	1026	252	**96.83**

Table 6. Confusion matrix for optimal values' C-SVM

	normal	Bg1L1	Bg1L2	Bg1R1	Bg1R2	Bg2L1	Bg2L2	Bg2R1	Bg2R2
normal	26	0	0	0	0	0	2	0	0
Bg1L1	0	27	0	0	0	0	0	1	0
Bg1L2	0	0	28	0	0	0	0	0	0
Bg1R1	0	1	0	26	0	0	0	1	0
Bg1R2	0	0	0	0	27	0	0	0	1
Bg2L1	0	0	0	0	0	26	0	2	0
Bg2L2	0	0	0	0	0	0	28	0	0
Bg2R1	0	0	0	0	0	0	0	28	0
Bg2R2	0	0	0	0	0	0	0	0	28

5 Conclusion

In this paper, an original process of fault detection and location experiments on single yaw damper has been presented. In order to obtain the potential useful information, dynamic performance indices, statistical parameters and wavelet energy moments have been extracted from different working conditions. Then, set of 434-dimensional feature vectors were obtained. Moreover, a hierarchical multi-class classification approach which was arranged into a BDT structure was applied in this special case. The structure of BDT was mainly composed of two parts: fault detection and fault localization. This algorithm takes advantage of the decision tree architecture and of the C-SVM. From the results presented above, this method outperformed other traditional methods. Last of all, we applied the BEE method to select the penalty factor C for linear C-SVM. Each C-SVM has its own optimal values which were listed in Table 4.

There are two representative conclusions which are as follows:

- The hierarchical multi-class classification approach is a meaningful method as it is consistent with basic two steps of fault diagnosis. With each nodes are parallel and independent, it is easier to further optimize. Especially in fault detection, the developed classifier based on OAO binary SVM can reduce the influence of deviation of hyper-plane.
- The BEE method for parameter selection is useful as it uses the prior knowledge and splitting operations are not required. From the results in section 4.3, the hierarchical multi-class classifier of optimal values improves by 5% compared with default values.

However, gradual faults occur frequently under practical operation. The slight fault of yaw damper will cause fatal accident directly. Fault diagnosis system should also detect the gradual fault of yaw damper. In the future, the fault quantification of yaw damper fault will be researched.

Acknowledgements

This work is supported by National Natural Science Foundation (NNSF) of China (61603316, 61433011).

References

[1] J.X. Liu, K.Y. Wang, Effect of yaw damper on locomotive riding comfort ability, Journal of Traffic and Transportation Engineering 6(2006) 1-4.

[2] Q.L. Liu, Y.H. Huang, Malfunction analysis and identification of EMU yaw damper, Machine Tool and Hydeaulics 44 (2016) 188-193.

[3] N. Qin, W.D. Jin, J. Huang, Feature extraction of high speed train bogie based on ensemble empirical mode decomposition and sample entropy, Journal of Southwest Jiaotong University 49(2014) 27-32.

[4] A. Alonso, J.G. Gimennez, E. Gomez, Yaw damper modeling and its influence on railway dynamic stability, Vehicle System Dynamics 49(2011) 1367-1387.

[5] H.Q. Patricia, J.B. Alonso, Review of automatic fault diagnosis systems using audio and vibration signals, IEEE Transactions on Systems, Man, and Cybernetics: Systems 44(2014) 642-652.

[6] H. Mechbal, J.S. Uribe, A probabilistic multi-class classifier for structural health monitoring, Mechanical Systems and Signal Processing 60(2015) 106-123.

[7] Q.H. Zhang, G.C. Zhao, L. Shu, Research of dimensionless index for fault diagnosis positioning based on EMD, Journal of Computers(Taiwan) 27(2016) 62-73.

[8] R. Jegadeeshwaran, V. Sugumaran, Fault diagnosis of automobile hydraulic brake system using statistical features and support vector machines, Mechanical Systems and Signal Processing 52(2015) 436-446.

[9] D.D. Yong, S. Bhowmik, An effective power quality classifier using wavelet transform and support vector machines, Expert Systems with Applications 42(2015) 6075-6081.

[10] G. Madzarov, D. Gjorgjevikj, I. Chorbev, A multi-class SVM classifier utilizing binary decision tree, Informatica 33(2009) 233-241.

[11] Y J. Lee, J.J Lee, Binary tree optimization using genetic algorithm for multiclass support vector machine, Expert Systems with Applications 42(2015) 3843-3851.

[12] J. Hernandez, L. Enrique, E.F. Morales, Multidimensional hierarchical classification, Expert Systems with Applications 41 (2014) 7671-7677.

[13] H.H Kim, J.Y. Choi, Hierarchical multi-class LAD based on OvA-binary tree using genetic algorithm, Expert Systems with Applications 42(2015) 8134-8145.

[14] L.A. Dalton, E.R. Dougherty, Bayesian minimum mean-square error estimation for classification error part I: definition and the Bayesian MMSE error estimator for discrete classification, IEEE Trans. Signal Process 59(2011) 115-129.

[15] L.A. Dalton, E.R. Dougherty, Bayesian minimum mean-square error estimation for classification error. Part II. The Bayesian MMSE error estimator for linear classification of Gaussian distributions, IEEE Trans. Signal Process 59(2011) 130-144.

[16] H. Huttunen, J. Tohka, Model selection for linear classifiers using Bayesian error estimation, Pattern Recognition 48(2015) 3739-3748.