JOURNAL OF IRON AND STEEL RESEARCH, INTERNATIONAL. 2011, 18(10): 17-23, 33

Multi-Class Classification Methods of Cost-Conscious LS-SVM for Fault Diagnosis of Blast Furnace

LIU Li-mei, WANG An-na, SHA Mo, ZHAO Feng-yun
(College of Information Science and Engineering, Northeastern University, Shenyang 110819, Liaoning, China)

Abstract: Aiming at the limitations of rapid fault diagnosis of blast furnace, a novel strategy based on cost-conscious least squares support vector machine (LS-SVM) is proposed to solve this problem. Firstly, modified discrete particle swarm optimization is applied to optimize the feature selection and the LS-SVM parameters. Secondly, cost-conscious formula is presented for fitness function and it contains in detail training time, recognition accuracy and the feature selection. The CLS-SVM algorithm is presented to increase the performance of the LS-SVM classifier. The new method can select the best fault features in much shorter time and have fewer support vectors and better generalization performance in the application of fault diagnosis of the blast furnace. Thirdly, a gradual change binary tree is established for blast furnace faults diagnosis. It is a multi-class classification method based on center-of-gravity formula distance of cluster. A gradual change classification percentage is used to select sample randomly. The proposed new method raises the speed of diagnosis, optimizes the classification accuracy and has good generalization ability for fault diagnosis of the application of blast furnace.

Key words: blast furnace; fault diagnosis; cost-conscious; LS-SVM; multi-class classification

Fault diagnosis of blast furnace is very important in the process industries. Gradually large-scale modern industry becomes integration and industrial processes become more complex. The number of measuring variables and requirements of operators are constantly changing. In this case, industrial process fault diagnosis is increasingly difficult and fault diagnosis is very urgent. And a serious blast furnace accident can take a few days or a week to make blast furnace normal. In this case, the loss of blast furnace iron production is thousands of tons and the consumption of coke is hundreds of tons more and the loss of economic benefits are millions of dollars. Some methods about fault diagnosis of blast furnace are used currently, such as neural network[1], fuzzy classification. A comparison between them is made[2]. Because fault diagnosis of blast furnace is a small sample problem, accuracy of support vector machine (SVM) is better than neural networks and fuzzy classification. Moreover, SVM is faster than the others. Therefore, it is imperative to solve fault diagnosis by SVM.

SVM is a learning approach for binary classification and is based on the theory of structural risk minimization[3-4]. Least squares support vector machine (LS-SVM) is a special case of a quadratic loss support vector machine model[5-8]. By its nature, fault diagnosis is a multi-class classification of fault data. Some multi-class classifiers are commonly used, such as one-against-one (OAO), one-against-all (OAA), pair-wise, directed acyclic graph (DAG) and decision-tree (DT) based SVM. But the OAO method constructs C_n^2 classifiers for n categories problems where each one is trained on data from two classes. It is cumbersome and waste of time. OAO method[2] was selected so that time is ignored. A model[9] was made by support vector machine ensemble. Various schemes[10] of OAA, pair-wise and directed acyclic graph SVM (DAGSVM) were compared, where OAA and DAGSVM were more suitable for practical use than pair-wise. But OAA will bring not to ensure the balance of the classification sample data in the course of training. This paper proposes an extended OAA and DT algorithm that is

Foundation Item: Item Sponsored by National Natural Science Foundation of China (60843007, 61050006)
Biography: LIU Li-mei(1979—), Female, Doctor Candidate; **E-mail:** liulimei_2006@163.com; **Received Date:** September 27, 2010

a gradual change binary tree so that this difficult issue is solved well. In this study, a new multi-class classification method based on cost-conscious LS-SVM is presented. The cost-conscious least squares support vector machine (CLS-SVM) algorithm is used for fault diagnosis of blast furnace. The binary tree multi-class classification based on center-of-gravity of distance formula of cluster is used to solve the faults diagnosis problem of blast furnace. How to ensure the balance of the classification sample data, the gradual change proportion of random sampling is presented so that it avoids the sample data imbalance because the "one-against-rest" classification brings to. A perfect gradual change binary tree is shown.

1 Least Squares Support Vector Machine

1.1 Binary LS-SVM classifier

Least squares support vector machine is proposed by Suykens[5]. To speed up SVM training, Suykens takes the quadratic function as the empirical risk function and substitutes equality constraints for the inequality ones. LS-SVM is the sum of the squared error and loss function as experience loss of a training set. In this version, one finds the solution by solving a set of linear equations instead of a time-consuming quadratic program (QP) for classical SVM[11]. This is due to the use of equality instead of inequality constraints in the problem formulation. Many pattern recognition and function approximation problems have been tackled with LS-SVM in the last decade[12-15].

Given samples $x_i \in R^n$, $i=1, \cdots, l$ and its classification is $y_i \in \{-1, 1\}$, $i=1, \cdots, l$, and l is the number of training samples. In two classes issues, x_i is input samples and y_i is the label corresponding input samples. Training set is

$$D=\{(x_i, y_i) \mid x_i \in R^n, y_i \in \{-1, 1\}, i=1, \cdots, l\} \quad (1)$$

When the data can not be separated linearly, kernel function is introduced. Function $\Phi(\cdot): R^n \rightarrow H$ is mapping samples of the input space to feature space. In the feature space H, inner product form is $\Phi(x_i) \cdot \Phi(x_j)$. If a function k can meet the condition of $k(x_i, x_j)=\Phi(x_i) \cdot \Phi(x_j)$, the inner product of high dimension space can be finished by the function in the original space. Decision function is the following form:

$$f(x)=\operatorname{sgn}[w^{\mathrm{T}} \cdot \Phi(x)+b] \quad (2)$$

In general, the optimization problem for LS-SVM is given following quadratic programming problem:

$$\min \frac{1}{2}\|w\|^2+\frac{\gamma}{2} \sum_{i=1}^{l} e_i^2$$
$$\text{s.t. } y_i[w \cdot \Phi(x_i)+b]=1-e, \quad i=1,2, \cdots, l \quad (3)$$

where, $\gamma>0$; e_i, $i=1, 2, \cdots, l$ are non-negative slack variables.

Referring to Suykens and Gestel's work[11], the decision function of LS-SVM can be obtained.

1.2 Multi-class LS-SVM classifiers based on binary tree

As the standard support vector machine supports only two-class classification, but in reality many problems often involve a multi-class classification. Currently, the main methods to solve multi-class classification problem are one-against-one, one-against-all, decision directed acyclic graph, the hierarchical tree, binary tree and so on. There is a lot of inseparable region for one-against-one and one-against-rest method. The hierarchical tree and binary tree method may solve this problem. The binary sub-tree given a reasonable definition or automatic calculation of artificial construction gets results after a few steps. It is faster than previous methods and has improved significantly. However, for the balance of the sample influence and a classification accuracy of the issue, the binary tree structure can be further optimized.

Between-class distance of cluster analysis is looked as a binary tree generation algorithm. The basic idea is to get as far as the other classes are related to the first division out of the class, then constructing the optimal hyper-plane should also have good generalization. Here is the calculation of the distance between the separation class and the others. Then the class corresponding to the maximum distance is divided out.

In order to make better use of training samples, the center-of-gravity method may be selected for between-class distance here. The basic idea of the center-of-gravity method is focused on that demand for the average value of all sample types as the center-of-gravity of class and the center-of-gravity distance between two classes is looked as the distance between two classes.

$$D_{i,j}=\|\overline{X^{(\omega_i)}}-\overline{X^{(\omega_j)}}\|;$$
$$\overline{X^{(\omega_i)}}=\frac{1}{N_i} \sum_{X \in \omega_i} X; \quad (4)$$
$$\overline{X^{(\omega_j)}}=\frac{1}{N_j} \sum_{X \in \omega_j} X;$$

where, $\|\cdot\|$ is Euclidean norm; and N_i, N_j are the number of samples about class ω_i, ω_j.

According to the above definition of the class distance, obviously there are $D_{i,i} = 0$, $D_{i,j} = D_{j,i}$. Since there is symmetry of the advantages of the distance, one half of the computation is saved at least[12].

2 Improved Particle Swarm Optimization Algorithms

Particle swarm optimization (PSO) is initialized with a group of random particles. Each particle flies in a d-dimensional space. The location of the i-th particle is expressed as $x_i = \{x_{i1}, x_{i2}, \cdots, x_{id}\}$. The location of the i-th particle to meet the best fitness value denotes $p_i = \{p_{i1}, p_{i2}, \cdots, p_{id}\}$. For a discrete problem expressed in a binary notation, a particle moves in a search space restricted to 0 or 1 on each dimension. In binary problem, updating a particle represents changes of a bit which should be in either state 1 or 0 and the velocity represents the probability of bit x_{id} taking the value 1 or 0. The best position of all particles in groups which can be shown $p_g = \{p_{g1}, p_{g2}, \cdots, p_{gd}\}$. Change the location of particle velocity as $v_i = \{v_{i1}, v_{i2}, \cdots, v_{id}\}$.

Each particle with two optimal values is constantly updated in each step. In both optimal values found, according to following Eqn. (5), the particles can adjust their speed and location[13-15]. In discrete particle swarm optimization, the equations of the position and the velocity of the particles are established as:

$$\begin{cases} v_{id}(t+1) = wv_{id}(t) + c_1\text{rand}()[p_{id} - x_{id}(t)] + \\ \qquad c_2\text{rand}()[p_{gd} - x_{id}(t)] \\ S[v_{id}(t+1)] = \tanh[(k+1)v_{id}(t+1)] \\ x_{id}(t+1) = 1 \text{ if rand}() < S[v_{id}(t+1)] \\ x_{id}(t+1) = 0 \text{ else} \end{cases} \tag{5}$$

where, rand() is a random number in interval (0, 1); c_1 and c_2 are two positive acceleration constants, which can be called the learning factor. The velocity v_{id} is generally conditioned by $v_{id} \in [-v_{max}, v_{max}]$ to prevent the particles from flying out of the solution area. In general, it $v_{max} = 6$ will be able to meet the condition.

On the basis of info-share mechanism of the PSO, by which the particle exchanges information only with the historical optimal solution of its own and the global historical optimal solution, Ref. [14] and Ref. [15] proposed respectively a new DPSO for variable selection. In order to simplify the problem, this paper proposes a modified DPSO algorithm and the random probability in Ref. [14] and Ref. [15] are replaced by the trisection point in the interval. It can make computing speed faster. The algorithm can be stated as following. If a selected random number rand() < 0.33 is satisfied, it may substitute x_{id}(old) for x_{id}(new). Otherwise, if they rand() $\geqslant$ 0.33 and v_{id} < 0.66 are satisfied, it may substitute p_{id} for x_{id}(new). While for the other situations, it may substitute p_{gd} for x_{id}(new). Where rand() are random numbers in the range of (0, 1); p_{id} is the optimal value of the particle itself; p_{gd} is the global optimal value; and v_{id} is the velocity which is the random number evenly distributed interval in [0, 1]. There should be a balance between the local and global search ability.

3 Multi-Class Classification Methods Based on Cost-Conscious LS-SVM

3.1 Cost-conscious variant model and optimization parameters

There are many ways solving fault diagnosis problem. In order to make better use of support vector machine to solve this problem, and take into account both accuracy and speed, an optimization of a LS-SVM concept model is presented. That is cost-conscious model. The cost-conscious model is a measure of LS-SVM. It is to estimate the overall cost that is including the SVM training accuracy, SVM training time and the characteristics of selected samples. Ref. [16] analyzed the cost-conscious on a number of support vector and kernel function selection of support vector machine. In order to facilitate the description of the problem, at first, some variables are introduced.

In LS-SVM, the relaxation factor e in SVM indicates, the error expectation in the classification process of the sample data. The penalty constants γ and the relaxation factor is in need of optimization. To simplify, we write the parameters γ, σ, e respectively as $t_1 = \gamma$, $t_2 = \sigma$, $t_3 = e$. Here σ is kernel parameter. The training model of SVM is $T = \{t_1, t_2, t_3\}$. The SVM parameter γ is limited in $[2^{-1}, 2^0, \cdots, 2^4]$ and σ is limited in $[2^0, 2^1, \cdots, 2^8]$ and the relaxation factor e satisfies $0 \leqslant e \leqslant 1$. The modified DPSO searches the optimal values of the SVM parameters in these areas.

Fault diagnosis of feature selection is to select variables from the overall classification results of some special variables[17-18]. The most representative characteristics of the feature of the data itself are selected, which is an optimization problem. Feature set has n feature vectors. The feature selection set is $F = \{f_1, f_2, \cdots, f_n\}$ where $f_i = 1$, $i = 1, \cdots, n$ denotes the i-th feature is selected and $f_i = 0$, $i = 1, \cdots, n$ shows the i-th feature is not done.

The classification speed of LS-SVM decreases, compared to standard SVM, especially for large scale problems, because the computation complexity of Eqn. (2) increases[19]. To reduce SV number, optimizing support vector set is imperative. Assume that $S = \{s_1, s_2, \cdots, s_l\}$ is a training set and has l vectors. It $s_j = 1$, $i = 1, \cdots, l$ denotes that the i-th sample is support vector and $s_i = 0$, $i = 1, \cdots, l$ shows that the i-th sample is not.

Combining the above parameters, a hybrid vector is (T, F, S). Firstly, let (T, F, S) jointly describe the training accuracy ratio of SVM, the feature selection and the number of support vectors. The joint optimization problem can be solved by the modified DPSO as well.

Secondly, a cost-conscious formula of cost-conscious LS-SVM (CLS-SVM) is defined as

$$c = f(A_{\text{test}-n}) \cdot \left[\frac{1}{l}\sum_{i=1}^{l} N(\alpha_i > 0)\right]^{-1} \cdot \left[\frac{1}{n}\sum_{i=1}^{l} N(f_i > 0)\right]^{-1} \tag{6}$$

where $f(A_{\text{test}-n})$ is the average value of the n-fold cross validations with data in the training set. All $\sum_{i=1}^{l} N(\alpha_i > 0)$ is respectively the number of the support vectors. Let $\frac{1}{l}\sum_{i=1}^{l} N(\alpha_i > 0)$ denote the percentage of the support vectors in the total samples and $\frac{1}{n}\sum_{i=1}^{l} N(f_i > 0)$ denote the percentage of the feature selection in the total feature. They are referred to as support vector percentage and feature percentage separately.

3.2 Fitness function based on cost-conscious model

Fitness function is used to evaluate the performance of each particle. It is the particle selected variables in the performance of fault diagnosis. Too many variables will increase complexity and real-time of the fault diagnosis. Therefore the unnecessary variables should be deleted to reduce dimension and improve accuracy rate of real-time fault diagnosis. Particles can choose a new fitness function $f(\text{new})$ as follows.

$$f(\text{new}) = f(A_{\text{test}-5}) \cdot \frac{l}{\sum_{i=1}^{l} N(\alpha_i > 0)} \cdot \frac{n}{\sum_{i=1}^{l} N(f_i > 0)} \tag{7}$$

where $f(A_{\text{test}-5})$ is the average accuracy rate of 5 fold cross validation for each particle in a training set.

3.3 Binary classification algorithm based on cost-conscious LS-SVM

Least squares support vector machine parameters and the kernel parameters affect the classification performance. The best values of these parameters can be adjusted by improved discrete particle swarm optimization method. A new binary classification algorithm of cost-conscious LS-SVM, that is called CLS-SVM, is proposed; and it conceives a new measure and can deal with the two-class classification problem better. The algorithm steps of CLS-SVM are as follows.

1) Initialization of the particle swarm. The parameters such as (T, F, S) should be involved in every particle. Each parameter is randomly valued in its limited area.

2) CLS-SVM is applied to train.

3) Calculating the fitness value of all particles.

4) Comparing the current fitness of a particle with its own historical maximal value, if its own maximal value is smaller then replacing it with the current fitness. Selecting the biggest fitness in the swarm is the global maximum. Refreshing the positions and velocities of the particle swarm according to the modified DPSO algorithm given in Section 2.

5) Judging whether the termination criterion is satisfied: if its answer is 'Yes', the iteration operation end, it goes to Step 6; if it is 'No', it goes to Step 1.

6) Obtaining the optimal parameters and feature selections.

7) CLS-SVM is training again.

3.4 Multi-class classification algorithm of cost-conscious LS-SVM based on gradual change binary tree

Because there is a problem of unbalanced data in OAA algorithm, this method will make low accuracy. To solve this problem, a new multi-class classification method: a gradual change binary tree is proposed. A gradual change classification percentage is used to select sample randomly and at any time to ensure the balance of the two classes of data. Here a clustering distance formula is introduced and it is Eqn. (4). According to the class distance formula, multi-class classification algorithm of the CLS-SVM based on gradual change binary tree is as follow that is to deal with fault diagnosis of the blast furnace.

1) According to the characteristics of blast furnace data, the normal state is selected for the first category, denoted n_1. In other words, the normal state is root of binary tree. Using distance Eqn. (4) and calculating the distances between the normal class and the fault state, denoted $D_{1,j}$ $(j = 2, \cdots, n)$. Select the maximum distance from $n-1$ values. If

two or more maximum distances appear at the same time, randomly a class is selected.

2) According to the maximum distance in step 1), the second class is determined, denoted n_2. It is seen as the next step of the node. Let calculate the distance between the second category and the other classes and be denoted $D_{2,j}(j=3, \cdots, n)$. Select the maximum distances from $n-2$ values. If two or more maximum distances appear at the same time, randomly a class is selected.

3) According to the maximum distance from the former class, the i-th class is determined, denoted n_i. It is seen as the next step of the node. Let calculate the i-th class and the distance between the other classes and be denoted $D_{i,j}(j=i+1, \cdots, n)$. Select the maximum distances from $n-i$ values. If two or more maximum distances appear at the same time, randomly a class is selected.

4) When it $j=n$ is satisfied, the algorithm stops. There is a complete binary tree and building order number $n_1, n_2, \cdots, n_n$. On the contrary, it goes to Step 3).

5) According to the generated binary tree, use binary CLS-SVM training algorithm to construct a binary tree and there is an optimal hyper-plane in each internal node. The n_1-th class of the sample set of the sample is positive and the other samples are negative class sample set. $\frac{1}{n-1}$ of the total number of negative class samples are randomly selected and use CLS-SVM algorithm to construct the binary classifier at the root node. Next, n_2-th class of the sample set of the sample is positive and the other samples $n_3, \cdots, n_n$ belong to negative class sample set. $\frac{1}{n-2}$ of the total number of negative class samples are randomly selected and use CLS-SVM algorithm to build the binary classifier at the second node. In turn, n_i-th class of the sample set of the sample is positive and the other samples $n_{i+1}, \cdots, n_n$ belong to negative class sample set. $\frac{1}{n-i}$ of the total number of negative class samples are randomly selected and use CLS-SVM algorithm to construct the binary classifier at the i-th node. At last, until all binary sub-classifier is trained to get the gradual change binary tree based on CLS-SVM multi-class classification model of CLS-SVM. In every step, binary classification is done for the balance of data. The training algorithm ends.

6) Input the test samples and obtain classification results.

7) Algorithm ends.

4 Simulation Experiments

4.1 Experimental background

Take a practical production data of a blast furnace for the background data, study is a different state of blast furnace testing and analysis, including normal and abnormal conditions. Abnormality includes heating, over-development of brim gas flow, hanging, cooling, chimney, lack of brim gas flow, low stock line, slip, such as eight common disorders and abnormalities. Simulation experiment of blast furnace use large-scale steel enterprise with the actual production data. The distribution of these data reflects the actual production situation. Experiment is completed on dual-core 3.16GHz Intel processor and 2GB of memory to operate on desktop computers.

BF state control is accomplished with many parameters. From the added raw materials, blast, fuel, iron ores are reduced to the final of the blast furnace iron production process during the process of a large number of process parameters. Once abnormal phenomenon of blast furnace appears, it would bring changes to some parameters[2,20]. In the production process, through the thermocouple temperature sensors and pressure sensors can obtain the temperature and pressure, and calculate the number of material per hour on the approval that was expected to material speed. The ratio of air volume and pressure is ventilation. Thus, selecting the state can get on the blast furnace and equipment mode of the parameters, such as wind, pressure, top pressure, differential pressure, permeability, the top temperature (includes four-point temperature), crossing temperature (including center and edge), feed rate, physical heat. Air volume and top pressure are basic parameters.

4.2 Experiment and analysis

The experiments are 540 groups of representative data. Experiment selected 40 normal samples and 320 fault samples. Each sample has eight feature vectors. From 540 samples, 360 samples are selected as training samples, including 40 normal samples, and 320 fault samples and each type of Fault of 40 samples. The remaining 180 samples are testing samples by multi-class classification algorithm of cost-conscious LS-SVM based on gradual change binary tree.

To facilitate illustration of the problem, class codes are used for the faults as shown in Table 1.

Table 1 Various types of data

Class code	Category	Number of training samples
a	Normal	40
b	Heating	40
c	Over-development of brim gas flow	40
d	Hanging	40
e	Cooling	40
f	Chimney	40
g	Lack of brim gas flow	40
h	Low stock line	40
i	Slip	40

According to the algorithm, a gradual change binary tree will appeare.

There is a classifier between a normal state and the other class b, c, d, e, f, g, h, i. Data selection is that class a has 40 samples, other class samples are randomly selected one-eighth of the total number of data. This ensures the balance of the data. Separate class a with other classes b, c, d, e, f, g, h, i.

Followed by the Eqn. (4), distance between class a and the other classes are calculated as shown in Table 2.

Table 2 Distance between class *a* and the others

Class a	Distance
$\Vert ab \Vert$	85.5270
$\Vert ac \Vert$	121.0628
$\Vert ad \Vert$	116.8920
$\Vert ae \Vert$	195.3408
$\Vert af \Vert$	722.5205 (maximum)
$\Vert ag \Vert$	228.3834
$\Vert ah \Vert$	248.0377
$\Vert ai \Vert$	214.0964

$\Vert af \Vert$ shows distance between class a and class f. The maximum value $\Vert af \Vert$ can be selected from Table 2 and f is determined to the second separation categories. There is a new classifier between class f and the others b, c, d, e, g, h, i. Class f has 40 samples. One-seventh of the total number of other samples is randomly selected so that balance classification is smooth. Then class f and other classes b, c, d, e, g, h and i will be separated.

So the distance values are as follows.

Table 3 shows the edge of failure and fault gas pipeline less than the maximum distance between the data, and other relatively distance of some faults data is small. It can be seen from Table 4 to Table 6 that class g, class h and class d are separated one by one by the same method.

Table 3 Distance between class *f* and the others *b*, *c*, *d*, *e*, *g*, *h*, *i*

Class f	Distance
$\Vert fb \Vert$	669.0005
$\Vert fc \Vert$	640.8393
$\Vert fd \Vert$	749.1102
$\Vert fe \Vert$	534.1640
$\Vert fg \Vert$	923.6914 (maximum)
$\Vert fh \Vert$	523.3655
$\Vert fi \Vert$	649.4150

Table 4 Distance between class *g* and the others *b*, *c*, *d*, *e*, *h*, *i*

Class g	Distance
$\Vert gb \Vert$	290.9531
$\Vert gc \Vert$	320.6519
$\Vert gd \Vert$	183.4765
$\Vert ge \Vert$	414.9158
$\Vert gh \Vert$	426.1085 (maximum)
$\Vert gi \Vert$	363.6350

Table 5 Distance between class *h* and the others *b*, *c*, *d*, *e*, *i*

Class h	Distance
$\Vert hb \Vert$	187.0131
$\Vert hc \Vert$	166.8468
$\Vert hd \Vert$	245.6747 (maximum)
$\Vert he \Vert$	141.4222
$\Vert hi \Vert$	215.2523

Table 6 Distance between class *d* and the others *b*, *c*, *e*, *i*

Class d	Distance
$\Vert db \Vert$	144.5635
$\Vert dc \Vert$	163.2622
$\Vert de \Vert$	255.7466 (maximum)
$\Vert di \Vert$	241.2061

As shown in Table 7, except for i class, the remaining two classes can be classified directly and the distance needs not to be mwasured.

Therefore, a complete binary tree forms what is valuable to solve fault diagnostics of blast furnace as shown in Fig. 1.

Table 7 Distance between class *e* and the others *b*, *c*, *i*

Class e	Distance
$\Vert eb \Vert$	143.7827
$\Vert ec \Vert$	141.5795
$\Vert ei \Vert$	217.0162 (maximum)

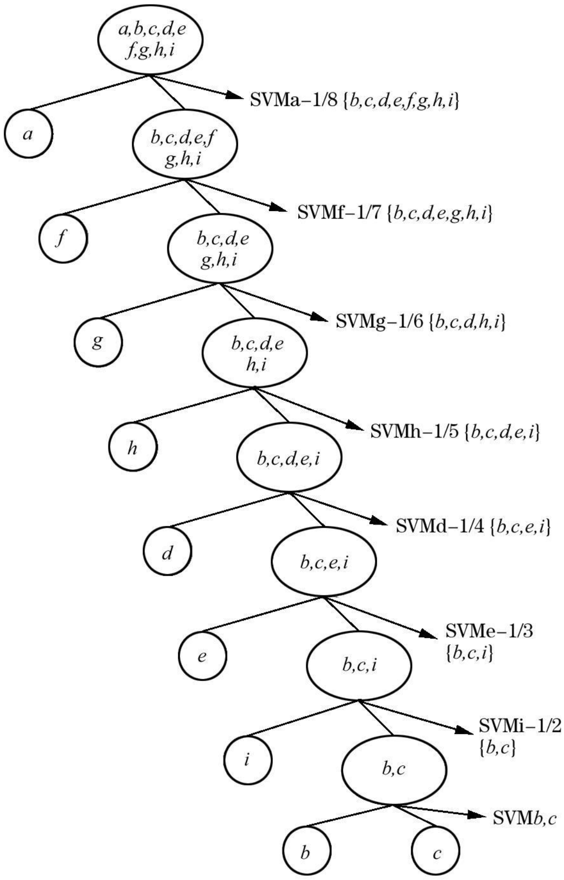

Fig 1 Gradual change binary tree based on the center of gravity distance formula

As seen from Fig. 1, 1/8 {b, c, d, e, f, g, h, i} means that the data contained one-eighth of the total number of eight categories samples. Similarly, 1/7 {b, c, d, e, g, h, i} represents the data contained one-seventh of the total number of seven classes samples. And so on and so forth.

Binary tree is constructed after the design is in accordance with the order followed by the binary classifier. Then the algorithm is trained on the data to obtain the best accuracy. After training the other data is tested to obtain diagnostic results as shown below.

There are LS-SVM and CLS-SVM methods for fault diagnosis of blast furnace and the results are as shown in Table 8. The first experiment applies the OAA algorithm for multi-class classification. The second experiment uses the gradual change binary tree to solve the multi-class classification. The results of experiments show that the second method has better or the same accuracy compared with the first method. While CLS-SVM takes less time than LS-SVM. Loss of relative time of the former is about 73% of the latter.

Table 8 Different diagnosis results under LS-SVM and CLS-SVM

Type	Error rate	
	LS-SVM	CLS-SVM
a	0.1124	0.1009
b	0.1214	0.1211
c	0.1013	0.1004
d	0.1214	0.1113
e	0.1113	0.1115
f	0.1264	0.1123
g	0.1271	0.1134
h	0.1143	0.1012
i	0.1145	0.1044

5 Conclusions

The proposed CLS-SVM algorithm presents a new generalization measure formula. CLS-SVM algorithm can determine the parameters of LS-SVM and has fewer number of support vectors and select the best features of the sample. Running time of CLS-SVM is relatively smaller than the others. CLS-SVM could be used as a faster and more reliable methodology. Gradual change binary tree is constructed by the center-of-gravity distance formula for multi-class faults. A gradual change proportion of random sampling method ensures the balance of the sample classification. This method can be used in the blast furnace diagnosis real-time. The results show that the proposed method has a better diagnosis accuracy compared with other methods and it has better generalization ability and diagnosis speed. This paper studies a multi-class classification with balance data. The next study will be focused on imbalanced data.

References:

[1] TANG Hong, LI Jing-min, YAO Bi-qiang, et al. Evaluation of Scheme Design of Blast Furnace Based on Artificial Neural Network [J]. Journal of Iron and Steel Research, International, 2008, 15(3): 1.

[2] WANG An-na, ZHANG Li-na, GAO Nan, et al. Fault Diagnosis of Blast Furnace Based on SVMs [C] // Proceedings of the 6th World Congress on Intelligent Control and Automation. Dalian: IEEE, 2006: 8.

[3] Vapnik, V N. The Nature of Statistical Learning Theory [M]. New York: Springer-Verlag, 1995.

[4] Joachims T. Estimating the Generalization Performance of a SVM Efficiently [C] // Proceedings of the 17th International Conference on Machine Learning. San Francisco: Morgan Kaufmann Publishers Inc, 2000: 431.

[5] Suykens A K, Gestel T V, Brabanter J D, et al. Least Squares Support Vector Machines [M]. Singapore: World Scientific Pub Co, 2002.

(Continued on Page 33)

References:

[1] Pinheiro C A, Samarasekera I V, Brimacombe J K. Mold Flux for Continuous Casting of Steel [J]. Iron and Steel Maker, 1995, 22(7): 41.

[2] Goldschmit M B, Gonzalez J C, Dvorkin E N. Finite Element Model for Analyzing Liquid Slag Development During Continuous of Round Bars [J]. Ironmaking and Steelmaking, 1993, 20(5): 379.

[3] Neumann F, Neal J, Pedroza M A, et al. Mold Fluxes in High Speed Thin Slab Casting [C] // Iron and Steel Society. Proceedings of 79th Steelmaking Conference. Pittsburgh: ISS, 1996: 249.

[4] Watanabe K, Suzuki M, Murakami K, et al. Development of Mold Powder for High Speed Casting of Middle Carbon Steel [C] // Iron and Steel Society. Proceedings of 79th Steelmaking Conference. Pittsburgh: ISS, 1996: 265.

[5] Lidefelt H, Hasselstrom P. Characterization of the Functional Properties of Mold Powders for Continuous Casting of Steel [C] // Metals Society. 4th International Iron and Steel Congress Proceedings. London: Metals Society, 1982: 101.

[6] LIU Cheng-jun, JIANG Mao-fa. Insulation Capacity of Mold Powder [J]. Journal of Iron and Steel Research, 2002, 14(3): 1 (in Chinese).

[7] Kishimoto M, Maeda M, Mori K, et al. Thermal Conductivity and Specific Heat of Metallurgical Slags [C] // Fine H A, Gaskell D R. Proceedings of 2nd International Symposium on Metallurgical Slags and Fluxes. Warrendale: Metallurgical Society of AIME, 1984: 891.

[8] Ponsford F H, Mills K C, Grievson P, et al. Physical and Thermal Properties of Aluminate Slag [C] // Institute of Metals. 3rd International Conference Molten Slags and Fluxes. Glasgow: Institute of Metals, 1988: 332.

[9] Nagata K, Susa M, Goto K S. Thermal Conductivities of Slags for Ironmaking and Steelmaking [J]. Tetsu-to-Hagane, 1983, 69: 1417 (in Japanese).

[10] Nagata K, Goto K S. Heat Conductivity and Mean Free Path of Phonons in Metallurgical Slags [C] // Fine H A, Gaskell D R. Proceedings of 2nd International Symposium on Metallurgical Slags and Fluxes. Warrendale: Metallurgical Society of AIME, 1984: 875.

[11] McDavid R M, Thomas B G. Flow and Thermal Behavior of the Top Surface Flux/Powder Layers in Continuous Casting Molds [J]. Metallurgical and Materials Transactions, 1996, 27B(4): 672.

[12] ZHANG Xian-zhuo. Metallurgical Transport Phenomena [M]. Beijing: Metallurgical Engineering Press, 1988 (in Chinese).

(Continued From Page 23)

[6] Guo X C, Yang J H, Wu C G, et al. A Novel LS-SVMs Hyper-Parameter Selection Based on Particle Swarm Optimization [J]. Neurocomputing, 2008, 71(16/17/18): 3211.

[7] ZENG Lian-ming, WU Xiang-bin, LIU Peng. Sample Reduction Strategy for SVM Large-Scale Training Data Set Using PSO [J]. Computer Science, 2009, 36(9): 215.

[8] Mitra V, Wang C J, Banerjee S. Text Classification: A Least Square Support Vector Machine Approach [J]. Appl Soft Comput, 2007(7): 908.

[9] TIAN Hui-xin, WANG An-na. A Novel Fault Diagnosis System for Blast Furnace Based on Support Vector Machine Ensemble [J]. ISIJ International, 2010, 50(5): 738.

[10] Hsu Chih-wei, Lin Chih-jen. A Comparison of Methods for Multiclass Support Vector Machines [J]. IEEE Transactions on Neural Networks, 2002, 13(2): 415.

[11] Suykens J A K, Lukas L, Vandewalle J. Sparse Approximation Using Least Squares Support Vector Machines [C] // Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS 2000). Geneva: Presses Polytechniques et Universitaires Romandes, 2000: 757.

[12] YANG Shu-ying. Pattern Recognition and Intelligent Computing [M]. 1st Beijing: Publishing House of Electronics Industry, 2008 (in Chinese).

[13] YUAN Sheng-fa, CHU Fu-lei. Fault Diagnostics Based on Particle Swarm Optimization and Support Vector Machines [J]. Mechanical Systems and Signal Processing, 2007, 21(24): 1787.

[14] Shen Q, Jianhui J. Modified Particle Swarm Optimisation Algorithm for Variable Selection in MLR and PLS Modeling: QSAR Studies of Antagonism of Antagonism of Angiotensin II Antagonists [J]. European Journal of Pharmaceutical Science, 2004, 22(2/3): 145.

[15] Kennedy J, Eberhart R. A Discrete Binary Version of the Particle Swarm Optimization Algorithm [C] // Proceedings of the IEEE International Conference on Systems, Mans and Cybernetics. Orlando: Insititute of Elactrical and Electronics Engineers, 1997: 4104.

[16] Mehmet Gonen, Ethem Alpaydin. Cost-Conscious Multiple Kernel Learning [J]. Pattern Recognition Letters, 2010, 31(9): 959.

[17] Polat K, Gunes S. A Novel Approach to Estimation of E. Coli Promoter Gene Sequences: Combining Feature Selection and Least Square Support Vector Machine (FS_LSSVM) [J]. Appl Math Comput, 2007, 190(2): 1574.

[18] Polat K, Kara S, Latifoglu F S. Pattern Detection of Atherosclerosis From Carotid Artery Doppler Signals Using Fuzzy Weighted Pre-Processing and Least Square Support Vector Machine (LSSVM) [J]. Ann Biomed Eng, 2007, 35(5): 724.

[19] TAO Shao-hui, CHEN De-zhao, ZHAO Wei-xiang. Fast Pruning Algorithm for Multi-Output LS-SVM and Its Application in Chemical [J]. Chemometrics and Intelligent Laboratory Systems, 2009, 96(1): 63.

[20] Chiang L H, Kotanchek M E, Kordon A K. Fault Diagnosis Based on Fisher Discriminant Analysis and Support Vector Machines [J]. Computers and Chemical Engineering, 2004, 28(8): 1389.