Strip Steel Surface Defect Classification Method Based on Enhanced Twin Support Vector Machine

Maoxiang CHU,^{1)*} Rongfen GONG¹⁾ and Anna WANG²⁾

1) School of Electronic and Information Engineering, University of Science and Technology Liaoning, Anshan, 114051 China.

2) College of Information Science and Engineering, Northeastern University, Shenyang, 110819 China.

(Received on July 28, 2013; accepted on September 19, 2013)

The strip steel surface defect classification belongs to multi-class classification. It demands high classification accuracy and efficiency. However, traditional methods are not fit for abnormal datasets, such as the large-scale, sparse, unbalanced and corrupted dataset. So a novel classification method is proposed in this paper based on enhanced twin support vector machine (TWSVM) and binary tree. According to the density information, the large-scale dataset is pruned, the sparse dataset is added with unlabeled samples, and TWSVM is improved to multi-density TWSVM (MDTWSVM) which has efficient successive over-relaxation (SOR) algorithm. Finally, MDTWSVM and binary tree are combined together to realize multi-class classification. Some experiments are done on the strip steel surface defect datasets with the proposed algorithm. Experimental results show that MDTWSVM has higher accuracy and efficiency than the other methods of multi-class classification for the strip steel surface defect.

KEY WORDS: strip steel; surface defect; multi-class classification; TWSVM; multi-density; SOR.

1. Introduction

The surface defect classification is an important step in the inspection techniques for the strip steel surface quality, because effective classification can not only improve the quality but also enhance the competitiveness of the strip steel products.¹⁾ As there are many kinds of defects during the producing and processing, such as scarring, dent, scratch, dirt, hole, damage and edge-crack, the strip steel surface defect classification belongs to multi-class classification problem. These defects should be completely classified on line so that the workers can deal with them as fast as possible. Moreover in the real classification condition, the defect samples are always unbalanced. That is to say, there are much more common defect samples than uncommon ones. And noise problems should be considered in real classification condition because noise samples can affect the performance of the classifier and even generate the classification error. So the main problem for the strip steel surface defect multi-class classification is how to get high-accuracy and high-efficiency classification results especially for large-scale, sparse, unbalanced and corrupted training datasets.

There are many methods for the strip steel surface defect classification,^{1–7)} while support vector machine (SVM) is proved to be the most effective classifier. In references^{4–7)} the standard or improved SVM is used to classify the strip steel surface defects, which gets better classification results compared to the other methods. However, there exists a con-

flict between the accuracy and efficiency in real application of SVM. So some novel SVMs^{8–11)} are appeared with the development of technology, such as fuzzy support vector machine, granular support vector machine, generalized eigenvalue proximal support vector machine (GEPSVM), and twin support vector machine (TWSVM). Among them the TWSVM is being widely used,^{12–14)} because not only it is approximately four times faster than the standard SVM but also it is fit for the cross plane dataset.

A novel method is proposed in this paper, which is more effective and accurate especially for large-scale, sparse, unbalanced and corrupted training datasets. This novel method is based on density information of training samples, where the large-scale training dataset is pruned, the sparse training dataset is added with the unlabeled samples, and the TWSVM is enhanced to multi-density TWSVM (MDTWSVM). Then the MDTWSVM, successive overrelaxation (SOR) technique and binary tree are combined to realize high-accuracy and high-efficiency multi-class classification for the strip steel surface defects.

This paper is structured as follows. In section 2 the TWSVM is introduced briefly. Section 3 focuses on extracting the density information for training datasets and processing large-scale and sparse training datasets. In section 4 the novel MDTWSVM are not only described, but also combined with SOR and binary tree to realize the multi-class classification. In section 5 some testing experiments are done and corresponding experimental results are described. Some conclusions are drawn in section 6.

* Corresponding author: E-mail: chu52_2004@163.com

DOI: <http://dx.doi.org/10.2355/isijinternational.54.119>

2. TWSVM

TWSVM is proposed by Javadeva in 2007, which is originated from GEPSVM. It generates two non-parallel planes such that each plane is closer to one of two classes and is as far as possible from the other.¹¹⁾ The TWSVM is four times faster than the standard SVM, for which it solves two smaller sized quadratic programming problems (QPPs) rather than one large QPP.¹¹⁾ TWSVM is described as the following two primal QPPs:

$$\begin{aligned} \min_{w_1, b_1} & \frac{1}{2} \sum_{i=1}^{m_1} (K(A_i', C')w_1 + b_1)^2 + c_1 \sum_{j=1}^{m_2} \xi_j \\ \text{s.t. } & K(B_j', C')w_1 + b_1 \leq -1 + \xi_j \quad \dots\dots\dots (1) \\ & \xi_j \geq 0, \quad j = 1, 2, \dots, m_2 \end{aligned}$$

$$\begin{aligned} \min_{w_2, b_2} & \frac{1}{2} \sum_{j=1}^{m_2} (K(B_j', C')w_2 + b_2)^2 + c_2 \sum_{i=1}^{m_1} \eta_i \\ \text{s.t. } & K(A_i', C')w_2 + b_2 \geq 1 - \eta_i \quad \dots\dots\dots (2) \\ & \eta_i \geq 0, \quad i = 1, 2, \dots, m_1 \end{aligned}$$

Where sample matrix $A = [A_1 \ A_2 \ \dots \ A_{m_1}]'$ belongs to class +1, sample matrix $B = [B_1 \ B_2 \ \dots \ B_{m_2}]'$ belongs to class -1, matrix $C = [A' \ B']'$, and K is any discretionary kernel. By introducing Lagrangian vectors, the Wolfe's dual of QPPs (1) and (2) can be given just like the following:

$$\begin{aligned} \min_{\alpha} & \frac{1}{2} \alpha' G (H' H)^{-1} G' \alpha - e_2' \alpha \\ \text{s.t. } & 0 \leq \alpha \leq c_1 \quad \dots\dots\dots (3) \end{aligned}$$

$$\begin{aligned} \min_{\beta} & \frac{1}{2} \beta' H (G' G)^{-1} H' \beta - e_1' \beta \\ \text{s.t. } & 0 \leq \beta \leq c_2 \quad \dots\dots\dots (4) \end{aligned}$$

Where $H = [K(A, C') \ e_1]$ and $G = [K(B, C') \ e_2]$. QPPs (3) and (4) determine two nonparallel kernel-generated surfaces $K(x', C')w_1 + b_1 = 0$, $K(x', C')w_2 + b_2 = 0$. By using the following linear kernel $K(x', C') = x' C'$, and defining the equations $u_1 = C'w_1$ and $u_2 = C'w_2$, the linear TWSVM classification hyper-planes $x'u_1 + b_1 = 0$ and $x'u_2 + b_2 = 0$ can be gotten. A new data sample x is assigned to the class +1 or -1 depending on which of the two hyper-planes lies closer to x in terms of perpendicular distance.¹¹⁾

3. Enhanced Methods for Training Dataset

TWSVM is more effective and accurate than the standard SVM. However, the classification results of TWSVM are not satisfied for the real time requirements in the strip steel surface defects multi-class classification. Moreover, the TWSVM can not tackle the large-scale, sparse, unbalanced and corrupted training datasets well. In order to solve these problems, the density information of every sample in each defect dataset is extracted and used. In detail, the large-scale dataset is pruned, and the sparse dataset is added with unlabeled samples according to that density information. All these methods are applied to TWSVM to enhance the performance of the TWSVM and realize optimal classification results.

3.1. Density Information of Training Dataset

Kernel density estimation (KDE) is a well developed statistic field.^{15,16)} It studies the feature of the data distribution from the data sample itself, which can be used in many real problems.¹⁶⁾ In this paper, KDE method is used to estimate density information of each sample in the training dataset. Consider a training dataset matrix $X = [X_1 \ X_2 \ \dots \ X_n]'$, and suppose that Ω_i is the nearest neighbor region with radius r and central point X_i , then density information of each sample in the training dataset can be defined following:

$$d_i = \sum_{\substack{X_t \in \Omega_i \\ t=1, 2, \dots, n}} \exp\left(-\frac{\|X_t - X_i\|^2}{(r/2)^2}\right) \quad i = 1, 2, \dots, n \quad \dots\dots\dots (5)$$

Where d_i reflects density information of the sample X_i , and r is set according to the real problem by the user. The bigger d_i , the more important the sample X_i in the training dataset. It can be seen from the Eq. (5), the density information of the sample X_i is determined by all samples in $\{X_t | X_t \in \Omega_i, t = 1, 2, \dots, n\}$. The closer the sample X_t is to the central sample X_i , the bigger the density contribution is to X_i . That is to say, the density information d_i is determined by the similarity between X_t and X_i in region Ω_i .

Density information d_i reflects the importance of sample X_i , which can be effectively used when the classifier deals with the training dataset. Take the training dataset with noise samples for example, the density value of noise sample is very small because the noise sample is isolated and the samples which are similar with it are very few. According to the density value, the noise samples of the training dataset can be deleted.

3.2. Pruning Samples for Training Dataset

For large-scale training dataset, it is necessary to prune samples for improving its classification efficiency. The pruning principle is not only to ensure the least effect on sample distribution of the training dataset, but also to reduce the number of the sample. In this paper the method that uses density information of large-scale dataset to guide the pruning process is proposed.

Consider a matrix $B = [B_1 \ B_2 \ \dots \ B_{m_2}]'$ belonging to large-scale training dataset, and suppose that Ω_l is the nearest neighbor region with radius r and the central point B_l , and d_l is the density information of sample B_l . Firstly, all density information $\{d_l | l = 1, 2, \dots, m_2\}$ is sorted in order. Then the sample B_l whose density information d_l is smaller than the threshold d is pruned from the large-scale training dataset. The pruned training dataset $\bar{B} = [\bar{B}_1 \ \bar{B}_2 \ \dots \ \bar{B}_{m'}]'$ is obtained. Finally, set the density information of the pruned samples from B to be 0 and the other densities to be d_j^- of $\bar{B}$. d_l and d_j^- are calculated just like the following:

$$\begin{aligned} d_j^- & \in \{d_l | I_l = 1, l = 1, 2, \dots, m_2\} \quad j = 1, 2, \dots, m^- \\ \text{where } I_l & = \begin{cases} 1 & d_l > d \\ 0 & d_l \leq d \end{cases} \quad l = 1, 2, \dots, m_2 \quad \dots\dots\dots (6) \\ d_l & = \sum_{B_t \in \Omega_l} \exp\left(-\frac{\|B_t - B_l\|^2}{(r/2)^2}\right) \quad l = 1, 2, \dots, m_2 \end{aligned}$$

Where, the threshold d is determined according to the real problem by user. It can be seen from the above formulas all

pruned samples provide the contribution to density information of $\bar{B}$. Both the classification efficiency and accuracy can be improved by using the density information d_j^- for large-scale training dataset.

3.3. Adding Unlabeled Samples for Training Dataset

For sparse training dataset, its complete sample distribution can not be reflected because there are not enough necessary samples in that dataset. To solve this problem, a novel method of selecting the samples from the vast of unlabeled samples and inserting these samples into the sparse training dataset is proposed in this paper. Unlabeled samples are added to the sparse training dataset by active learning¹⁷⁾ and their density information is used to guide the process.

Consider a matrix $A = [A_1 A_2 \dots A_{m_1}]'$ belonging to sparse training dataset, and suppose that Ω_i is the nearest neighbor region with radius r and central point A_i , and d_i is the density information of sample A_i . Firstly, the unlabeled sample X_p in region Ω_i will be selected. Then the sample X_p becomes a labeled sample of the training dataset, which can enhance the density of the sample A_i . Secondly, the center A_c of the training set will be calculated. Then a sphere region Ω_c with the central point A_c and radius R will be constructed. The unlabeled sample X_q that is not in the region Ω_i but in the region Ω_c will be selected. The sample X_q becomes a labeled sample of the training dataset. Then insert the new labeled sample X_q into the training dataset A to obtain the new training dataset $\bar{A} = [\bar{A}_1 \bar{A}_2 \dots \bar{A}_{m_1} \bar{A}_{m_1+1} \bar{A}_{m_1+2} \dots \bar{A}_{m^+}]' = [A_1 A_2 \dots A_{m_1} X_1 X_2 \dots X_{m^+-m_1}]'$. Finally, reconstruct Ω_i and calculate the sample density d_i^+ for the new training dataset $\bar{A}$. The parameters, such as A_c , R , d_i and d_i^+ , are calculated just like the following:

$$A_c = \frac{1}{m_1} \sum_{i=1}^{m_1} A_i, \quad R = \max_{i=1,2,\dots,m_1} \|A_i - A_c\| - r \quad \dots (7)$$

$$d_i^+ = \begin{cases} d_i + \sum_{X_p \in \Omega_i} \exp\left(-\frac{\|X_p - A_i\|^2}{(r/2)^2}\right) & i=1,2,\dots,m_1 \\ \sum_{X_t \in \Omega_i} \exp\left(-\frac{\|X_t - X_{i-m_1}\|^2}{(r/2)^2}\right) \\ + \sum_{A_t \in \Omega_i} \exp\left(-\frac{\|A_t - X_{i-m_1}\|^2}{(r/2)^2}\right) & i=m_1+1, m_1+2, \dots, m^+ \end{cases}$$

where $d_i = \sum_{A_i \in \Omega_i} \exp\left(-\frac{\|A_i - A_i\|^2}{(r/2)^2}\right) \quad i=1,2,\dots,m_1$

..... (8)

The sample distribution of sparse training dataset will be improved by adding unlabeled samples. It can be seen clearly from the above formulas that these added unlabeled samples enhanced the density of not only the original samples but also the integral samples of training dataset. The density information d_i^+ can be used to improve classification accuracy for sparse training dataset.

4. Multi-density TWSVM and Multi-class Classification Algorithm

4.1. Multi-density TWSVM

In order to get high-accuracy and high-efficiency multi-class classification results especially for large-scale, sparse, unbalanced and corrupted training datasets, TWSVM is improved to be multi-density TWSVM based on enhanced methods in section 3.

Consider the following binary classification problem. Suppose $A = [A_1 A_2 \dots A_{m_1}]'$ is a labeled sparse training dataset, which belongs to the class +1. It becomes $\bar{A} = [\bar{A}_1 \bar{A}_2 \dots \bar{A}_{m^+}]'$ after being added by the new labeled samples. And suppose $B = [B_1 B_2 \dots B_{m_2}]'$ is a labeled large-scale training dataset, which belongs to the class -1. And the matrix $\bar{B} = [\bar{B}_1 \bar{B}_2 \dots \bar{B}_{m^-}]'$ is a pruned training dataset on the base of B . Based on what mentioned above, MDTWSVM is described as the following two primal QPPs:

$$\min_{w_1, b_1} \frac{1}{2} \sum_{i=1}^{m^+} d_i^+ (K(\bar{A}_i', C') w_1 + b_1)^2 + c_1 \sum_{j=1}^{m^-} d_j^- \eta_j$$

s.t. $K(\bar{B}_j', C') w_1 + b_1 \leq -1 + \eta_j, \quad \eta_j \geq 0, \quad j=1,2,\dots,m^-$

..... (9)

$$\min_{w_2, b_2} \frac{1}{2} \sum_{j=1}^{m^-} d_j^- (K(\bar{B}_j', C') w_2 + b_2)^2 + c_2 \left(\sum_{i=1}^{m^+} d_i^+ \xi_i \right)$$

s.t. $K(\bar{A}_i', C') w_2 + b_2 \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i=1,2,\dots,m^+$

..... (10)

Where $C' = [\bar{A}' \quad \bar{B}']$, and c_1 and c_2 represent penalty factors. d_i^+ is density information of sample in training dataset $\bar{A}$. d_j^- is density information of sample in training dataset $\bar{B}$. d_j^- and d_i^+ can be obtained by using Eqs. (6) and (8). Optimizing the primal QPP (9) of MDTWSVM tends to make the hyper-plane close to class +1 and far to class -1 at a distance of at least 1. However, for class +1, the samples with big density values are possible to be near the hyper-plane and the samples with small density values have little effect on the hyper-plane because of added density parameters. At the same time, for the class -1, the introduction of density parameters makes it possible for the hyper-plane to be at a distance of at least 1 from the samples with big density values. Multi-density information of the training samples can effectively reduce effect of noise samples and improve the classification accuracy. The primal QPP (10) of MDTWSVM has similar results.

The Lagrangian corresponding to the primal QPP (9) of MDTWSVM is given:

$$L(w_1, b_1, \alpha_j, \beta_j, \eta_j) = \frac{1}{2} \sum_{i=1}^{m^+} d_i^+ (K(\bar{A}_i', C') w_1 + b_1)^2 + c_1 \sum_{j=1}^{m^-} d_j^- \eta_j$$

$$+ \sum_{j=1}^{m^-} \alpha_j (K(\bar{B}_j', C') w_1 + b_1 + 1 - \eta_j) - \sum_{j=1}^{m^-} \beta_j \eta_j$$

..... (11)

Using Karush-Kuhn-Tucker conditions and a serial similar process to TWSVM in reference,¹¹⁾ the Wolfe dual of the primal QPP (9) of MDTWSVM is obtained:

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \alpha' G (H' D^+ H)^{-1} G' \alpha - e_2' \alpha \\ \text{s.t.} \quad & 0 \leq \alpha \leq c_1 D^- e_2 \end{aligned} \quad \dots\dots\dots (12)$$

Where $\alpha = (\alpha_1 \ \alpha_2 \ \dots \ \alpha_m)'$, $H = [K(\bar{A}, C') \ e_1]$ and $G = [K(\bar{B}, C') \ e_2]$. D^+ is a diagonal matrix with the diagonal element d_i^+ . D^- is a diagonal matrix with the diagonal element d_j^- . The corresponding hyper-plane $K(x', C')w_1 + b_1 = 0$ can be obtained through the primal QPP (12) of MDTWSVM. And the parameters w_1 and b_1 can be calculated with the following formula:

$$\begin{bmatrix} w_1 \\ b_1 \end{bmatrix} = -(H' D^+ H)^{-1} G' \alpha \quad \dots\dots\dots (13)$$

In a similar way, the Wolfe's dual of the primal QPP (10) of MDTWSVM can be obtained, and the hyper-plane $K(x', C')w_1 + b_1 = 0$ is solved by following:

$$\begin{aligned} \min_{\mu} \quad & \frac{1}{2} \mu' H (G' D^- G)^{-1} H' \mu - e_1' \mu \\ \text{s.t.} \quad & 0 \leq \mu \leq c_2 D^+ e_1 \end{aligned} \quad \dots\dots\dots (14)$$

$$\begin{bmatrix} w_2 \\ b_2 \end{bmatrix} = (G' D^- G)^{-1} H' \mu \quad \dots\dots\dots (15)$$

Where $\mu = (\eta_1 \ \eta_2 \ \dots \ \eta_m)'$. A new data sample x is assigned to the class +1 or -1 depending on which of the two hyper-planes lies closer to x in terms of perpendicular distance.

$$i = \arg \min_{k=1,2} \frac{|K(x', C')w_k + b_k|}{\|w_k\|}, \quad x \in \begin{cases} \text{class } +1 & \text{if } i=1 \\ \text{class } -1 & \text{if } i=2 \end{cases} \quad \dots\dots (16)$$

4.2. SOR Technique for MDTWSVM

In MDTWSVM, there are two QPPs (12) and (14) to be solved. The primal QPP (12) is rewritten in the following form by defining $U = G(H'D^+H)^{-1}G'$, $c = c_1 D^- e_2$ and $\varphi = \alpha$:

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \varphi' U \varphi - e' \varphi \\ \text{s.t.} \quad & 0 \leq \varphi \leq c \end{aligned} \quad \dots\dots\dots (17)$$

SOR for quadratic programs is a high effective iterative method, which is considerably faster than other methods. It has been proved that this algorithm converges linearly to a solution.¹⁸⁾ In reference,¹⁸⁾ the SOR technique has been used effectively to improve the efficiency of solving QPP. The paper also employs the SOR technique to solve the primal QPP (17). The iterative equations of the SOR for the QPP (17) are just like the following:

$$\varphi_j^{k+1} = \varphi_j^k - \rho E_{jj}^{-1} \left(\sum_{l=1}^{j-1} U_{jl} \varphi_l^{k+1} + \sum_{l=j}^m U_{jl} \varphi_l^k - 1 \right) \quad j=1,2,\dots,m^- \quad \dots\dots\dots (18)$$

Where $\rho \in (0,2)$. E_{jj} is the diagonal element of matrix U . k is the iterative times. It can be seen from the Eq. (18) that the iterative process begins with an initial φ^0 , and then φ^{k+1} is calculated from φ^k , until $\|\varphi^{k+1} - \varphi^k\|$ is less than some prescribed tolerance ε . During the iterative process, φ^{k+1} is restricted within the interval $[0, c]$.

The primal QPP (14) is also rewritten as QPP (17) by

defining $U = H(G'D^-G)^{-1}H'$, $c = c_2 D^+ e_1$ and $\varphi = \mu$, which can be calculated with the Eq. (18).

4.3. Multi-class Classification Algorithm

The strip steel surface defect classification belongs to multi-class classification problem. There are many multi-class classification methods based on binary classifier, such as one-against-one, one-against-rest, decision directed acyclic graph and binary tree.¹⁹⁾ And binary tree is most widely used. The basic binary tree model for seven classes is shown in Fig. 1.

It can be seen from Fig. 1 that the binary tree improves the classification efficiency by constructing 6 classifiers for seven classes. However, the training dataset may be unbalanced for every binary classifier. So it is necessary to solve the unbalanced problem when the binary tree is used to realize multi-class classification. In this paper the binary tree and binary classifier MDTWSVM are combined. In MDTWSVM, the class +1 represents the sparse training dataset A and the class -1 represents the large-scale training dataset B . When every node classifier of binary tree is constructed, it should be confirmed to be class +1 or class -1 according to the number of samples in training dataset. Suppose there are N classes for strip steel surface defects in training dataset, the multi-class classification algorithm by combining binary tree and binary MDTWSVM is obtained. The algorithm steps are just like the following:

Step 1: analyze training datasets of N classes and decide sparse and large-scale training datasets. Then select rational parameters, such as radius r and pruning threshold d . And add unlabeled samples to the sparse training dataset and prune samples for the large-scale training dataset to get new training datasets based on enhanced methods in section 3.

Step 2: select one class of samples as training dataset 1, and the others as training dataset 2 for the current node of binary tree. Then determine the sparse training dataset A of class +1 and the large-scale training dataset B of class -1 according to the numbers of the training datasets 1 and 2.

Step 3: obtain new added training dataset $\bar{A}$ from A and the new pruned training dataset $\bar{B}$ from B according to the information of step 1.

Step 4: calculate density information d_j^- and d_i^+ for training datasets $\bar{B}$ and $\bar{A}$ according to Eqs. (6) and (8) respectively.

Step 5: select rational parameters c_1 , c_2 and kernel function K .

Fig. 1. Binary tree for seven classes.

Fig. 2. Images of seven surface defects.

Step 6: define G, H, D^+, D^-, U and select parameter ρ . And calculate Lagrangian vectors α and μ according to Eq. (18).

Step 7: calculate parameters of the two nonparallel hyperplanes with Eqs. (13) and (15) to obtain the classifier for this node.

Step 8: repeat steps from 2 to 7 until obtain all the nodes classifiers of binary tree.

Step 9: According to the binary tree, calculate every node classifier by Eq. (16) until a new defect sample falls into one of N classes.

5. Experiments

5.1. Description of Strip Steel Surface Defect Datasets

It is important to obtain a complete database of strip steel surface defect before classifying. It is well known that computer vision technology has been widely used in the strip steel surface defect detection based on charged coupled device (CCD). All experimental images of the strip steel surface defect are obtained by using CCD surface detection system of national large steel plants in this paper. Four classes of defect images with large-scale samples and three classes with sparse samples are selected. These defects are scarring, dent, scratch, dirt, hole, damage and edge-crack, which are shown in Fig. 2.

A defect sample is a multi-dimensional vector, which is obtained through extracting features of defect image. These features have immediate impact on the classification accuracy and efficiency. These defect features including geometry, gray, projection, texture and frequency-domain features²⁰⁾ are often extracted. And each of these features involves multi-dimensional information. In this paper there are 64 defect features for each defect image, which compose a 64-dimensional defect sample. In order to ensure the efficiency, manifold learning algorithm²¹⁾ is used to reduce the dimension to be 34. So each sample is a 34-dimensional vector.

In this paper, 2 330 images are selected with seven classes of defect images, which become 34-dimensional samples through feature extraction and dimensionality reduction. These samples are provided to do testing classification experiments. Then they are divided into training dataset and testing dataset randomly. The dividing results are shown in Table 1.

5.2. Parameters Analysis for MDTWSVM

As discussed in section 3 and 4, six parameters of

Table 1. Different surface defect data samples.

Defect type	Number of training samples	Number of testing samples
Scarring	477	53
Dent	468	52
Scratch	450	50
Dirt	459	51
Hole	72	18
Damage	81	19
Edge-crack	63	17

MDTWSVM algorithm, such as r, d, ρ, c_1, c_2 and K , will affect the performance of multi-class classification algorithm. It is necessary to consider how to select the optimal values for these parameters when MDTWSVM algorithm is used to realize the strip steel surface defect multi-class classification. Parameter r is the radius of nearest neighbor region that affects density information of all samples in training dataset. The smaller r , the more independent the sample, and vice versa. Parameter d is pruning threshold for pruning samples, which affects pruning ratio. Too small d is not satisfied the demand of pruning ratio, and too large d affects the classification accuracy. Parameter ρ is the slack factor of the SOR algorithm, which affects the convergence. When penalty factors c_1 and c_2 are optimal the performance of the MDTWSVM algorithm is satisfied. Radial basis function (RBF) is often used as kernel function K of MDTWSVM. For RBF, the corresponding parameter is the kernel radius δ that directly affects the classification accuracy of MDTWSVM.

In this paper, all training datasets are firstly normalized and then six parameters are determined. In general, the radius r and the threshold d should be selected from 0.2 to 0.4. Slack factor ρ should be defined in the interval [0,2]. Penalty factors c_1 and c_2 are obtained from 2^{-10} to 2^5 . And kernel radius δ is from 2^{-10} to 2^4 . These parameters can be optimized by using some usual methods, such as recursive exhaustion technique, gradient descent method and genetic algorithm. The tenfold cross-validation method^{16,18)} is used to obtain these optimal parameters in this paper.

5.3. Experiments and Results Analysis

Four experiments need to be done to testify the efficiency of multi-class classification algorithm with MDTWSVM and binary tree for the strip steel surface defect. Firstly, 3 experiments are used to test the effects on which the added sparse sample has, the pruned large-scale sample has, and the density information algorithm has respectively. And the last one is to make comparisons among 3 different classifiers. All experimental results are shown in order in the following content.

Firstly, corresponding to different pruning ratio for large-scale training dataset, different testing results are shown in Table 2. To mention that the pruning ratio is the ratio of the number of pruned data samples to that of the original data samples. From Table 2, it is clear that the larger the ratio, the shorter the training and testing time, but the lower the classification accuracy. Pruning ratio should be appropriately selected by considering both time and accuracy in real

Table 2. Testing results with different pruning ratio.

Pruning ratio	Training time (s)	Testing time (s)	Accuracy (%)
0.00	2.6447	0.2753	95.38
0.30	1.4555	0.2251	95.00
0.40	1.1243	0.0902	94.62
0.50	0.8713	0.0920	89.23
0.70	0.1340	0.0778	76.92
0.80	0.0618	0.0362	61.15

Table 3. Testing results with different adding ratio.

Adding ratio	Training time (s)	Testing time (s)	Accuracy (%)
0.00	1.1243	0.0902	94.62
0.30	1.1865	0.0996	95.00
0.50	1.2213	0.1022	96.15
0.60	1.2623	0.1068	96.92
0.70	1.3040	0.1109	96.54
0.90	1.3618	0.1141	94.23

Table 4. Classification accuracy of different defects.

Defect type	Accuracy (%)	
	Not using MD	Using MD
Scarring	94.34	98.11
Dent	96.15	100
Scratch	90.00	96.00
Dirt	90.20	96.08
Hole	94.45	94.45
Damage	89.47	94.74
Edge-crack	88.24	94.12

Table 5. Testing results of three classifiers.

Classifier	Training time(s)	Testing time (s)	Accuracy (%)
SVM	91.1512	1.1886	92.31
TWSVM	11.3707	0.6881	93.85
MDTWSVM	1.2623	0.1068	96.92

classification problem. So the pruning ratio is determined as 0.40 in this paper.

Secondly, corresponding to different adding ratio for sparse training dataset, different testing results are shown in **Table 3**. To mention that the adding ratio is the ratio of the number of added unlabeled samples to that of the original samples. It is necessary to point out that the experiment is made when the pruning ratio is 0.4 for large-scale training dataset. It can be seen from Table 3 that the larger the adding ratio, the longer the training and testing time. However, classification accuracy is firstly high then low as the ratio becomes large. So the adding ratio is determined as 0.60 in this paper.

Thirdly, in order to test the efficiency of the density information parameter in MDTWSVM, classification accuracy is tested on training dataset with multi-density (MD) parameter or not respectively. It can be seen from **Table 4** that the

accuracy is much higher in using MD than not using MD. So combining adding unlabeled samples for sparse training dataset, pruning samples for large-scale training dataset and using multi-density information, can enhance classification accuracy effectively.

Finally, in order to verify the efficiency of the algorithm proposed in this paper, comparative experiments are made with MDTWSVM, standard SVM and TWSVM. In order to be fair, the same binary tree is used in these different methods. The final 3 results are shown in **Table 5**. It is clear that the MDTWSVM proposed in this paper is much better than the others on time cost and classification accuracy.

6. Conclusions

A novel multi-class classification algorithm for the strip steel surface defects is propose based on density information of training dataset in this paper. Density information is used to guide these processes of pruning large-scale training dataset, adding unlabeled samples for sparse training dataset and improving TWSVM to MDTWSVM. And the SOR method is used to resolve the QPPs of MDTWSVM. The novel method proposed in this paper is fit for multi-class, large-scale, sparse, unbalanced and noised sample dataset. Finally, the results of experiments made on the strip steel surface defect dataset show that MDTWSVM combining binary tree is of perfect classification accuracy and efficiency.

REFERENCES

- 1) L. A. O. Martins, F. L. C. Pádua and P. E. M. Almeida: Proc. of 36th Annual Conf. on IEEE Industrial Electronics Society, IEEE, Piscataway, NJ, (2010), 1081.
- 2) S. R. Aghdam, E. Amid and M. F. Imani: Proc. of 7th IEEE Conf. on Industrial Electronics and Applications, IEEE, Piscataway, NJ, (2012), 1447.
- 3) M. Yazdchi, M. Yazdi and A. G. Mahyari: Proc. of Int. Conf. on Digital Image Processing, IEEE, Piscataway, NJ, (2009), 346.
- 4) K. Agarwal, R. Shivpuri, Y. J. Zhu, T. S. Chang and H. Huang: *Expert Syst. Appl.*, **38** (2011), 7251.
- 5) Y. J. Chen, L. Chen, X. M. Liu, S. Ding and H. Zhang: Practical Applications of Intelligent Systems, Vol. 124, Springer-Verlag, Berlin Heidelberg, (2012), 185.
- 6) Y. W. Yu, G. F. Yin and L. Q. Du: *Adv. Mater. Res.*, **217–218** (2011), 336.
- 7) E. Amid, S. R. Aghdam and H. Amindavar: *World Acad. Sci., Eng. Technol.*, **67** (2012), 1303.
- 8) C. F. Lin and S. D. Wang: *IEEE Trans. Neural Netw.*, **13** (2002), 464.
- 9) Y. C. Tang, B. Jin, Y. Sun and Y. Q. Zhang: Proc. of 2004 IEEE Symp. on Computational Intelligence in Bioinformatics and Computational Biology, IEEE, Piscataway, NJ, (2004), 73.
- 10) O. L. Mangasarian and E. W. Wild: *IEEE Trans. Pattern Anal. Mach. Intell.*, **28** (2006), 69.
- 11) Jayadeva, R. Khemchandani and S. Chandra: *IEEE Trans. Pattern Anal. Mach. Intell.*, **29** (2007), 905.
- 12) X. J. Ding, G. L. Zhang, Y. Z. Ke, B. L. Ma and Z. C. Li: Proc. of Int. Symp. on Information Science and Engineering, Vol. 1, IEEE, Piscataway, NJ, (2008), 560.
- 13) H. H. Cong, C. F. Yang and X. R. Pu: Proc. of IEEE Conf. Robotics, Automation and Mechatronics, IEEE, Piscataway, NJ, (2008), 348.
- 14) G. R. Naik, D. K. Kumar and Jayadeva: *IEEE Trans. Inf. Technol. Biomed.*, Vol. 14, IEEE, Piscataway, NJ, (2010), 301.
- 15) J. L. Li and H. G. Fu: *Control Decision*, **25** (2010), 507 (in Chinese).
- 16) X. J. Peng and D. Xu: *Neurocomputing*, **99** (2013), 134.
- 17) L. M. Liu, A. N. Wang, M. Sha, X. Y. Sun and Y. L. Li: *ISIJ Int.*, **51** (2011), 1474.
- 18) Y. H. Shao, C. H. Zhang, X. B. Wang and N. Y. Deng: *IEEE Trans. Neural Netw.*, **22** (2011), 962.
- 19) L. M. Liu, A. N. Wang, M. Sha and F. Y. Zhao: *J. Iron Steel Res. Int.*, **18** (2011), 17.
- 20) Y. Zhang, W. W. Liu, Z. T. Xing and Y. H. Yan: *J. Northeastern Uni. (Natural Science)*, **33** (2012), 267 (in Chinese).
- 21) M. Wang, H. Qiao and B. Zhang: *IEEE Trans. Intell. Transp. Syst.*, **12** (2011), 1195.