The Classification Algorithm of Defects in Weld Image Based on Weighted SVM

ZHANG xiao-guang[1,2] XU jian-jian[1] REN shi-jin[3]
1. Institute of Applied Physics Nanjing university, Nanjing 210093, P.R. China
2.College of Mechatronic Engineering of China University of Mining and Technology, Xuzhou 221008, P.R.China
3. Zhejiang University National Lab of Industrial Control Technology, Hangzhou 310027, P.R.China
E-mail: doctorzxg@163.com

ABSTRACT: *This paper firstly analyzes the Classification principle of SVM and indicates that SVM can not obtain favorable classification ability when the numbers of all classes of samples vary greatly. The algorithm of weighted SVM is put forward based on the analysis of the reason why the classification inclination comes into being, which can compensate the effect of the uneven class sizes and advance the classification accuracy of the smaller sample size. The experimental results of defect recognition in weld image show that this algorithm can improve the accuracy of small class effectively.*
Keywords: *Weighted; Support Vector Machine (SVM); Uneven class sizes; Classification accuracy; Weld defects, Pattern recognition*

1. INTRODUCTION

Welded structures often have to be tested nondestructively, since weld failure can be catastrophic, some significant components such as in pressure vessels, power plants and aeronautic equipment must be tested. When defects in materials and weld are found using radiographic inspection, the size and the class of defect must be determined. Operators depend on qualitative and quantitative analysis to make judgment in nondestructive testing. Nowadays radiographic detection relies on human experts to perform interpretation of images, which leads to operating troublesome and uncertainty judging result. With the help of computer assessing the quality of weld images make great progress in radiographic inspection. In the last 30 years, the traditional classifiers were used to recognize weld defects and little success was made [1][2]. Recently application of neural network (NN) to defects recognition can improve accuracy greatly[3][4].But to the neural network the structure of its hidden layers does not follow certain rules and is easy to trap local minimum, so these problems decrease the actual application of neural network to a certain extent. The algorithm of Support Vector Machine (SVM) can solve them effectively.

According to the principle of structure risk minimum, Vapnik et al [4] proposed a new statistic learning methodology-SVM, which can increase the generality of learning machine. It employs some limited samples to train and can get small errors to independent testing set. But the accuracy of SVM (generalization ability) is related to the number of support vectors. Under the certain training accuracy, the less is the number of support vector, the better is the generalization ability. However, the causes that lead to excessive support vectors of SVM are so many. Since the number of defect class samples is different, it is easy to result in recognition mistake. This paper, according to the characteristics of weld defects, presents a recognition method of weld defects based on weighted SVM.

2. THE EFFECTS THAT SVM DOES ON THE CLASSIFICATION WHEN DEALING WITH UNEVEN NUMBERS OF SAMPLES

Generally speaking, training vectors $x_i \in R^{(n)}, i = 1, \cdots, l$ belong to two classes, $y_i \in \{1,-1\}$. Thus the objective function of SVM is[5,6]:

$$\min \quad \frac{1}{2}\|w\|^2 + C\left(-\nu\rho + \frac{1}{l}\sum_{i=1}^{l}\xi_i\right)$$
$$s.t. \quad \begin{cases} y_i(w^T\phi(x_i)+b) \ge \rho - \xi_i \\ \xi_i \ge 0, i = 1, \cdots, l \\ \rho \ge 0 \end{cases} \tag{1}$$

In the formula above, s.t.stands for subject to and is the meaning of constrain. In the case of linear division, exist (w,b) which makes $w \cdot x_i + b > 0, \forall x_i \in Class\,1$, and $w \cdot x_i + b < 0, \forall x_i \in Class\,2$

The pairing form of the objective function of (1) formula is as follows:

$$\min \quad \sum_{i,j=1}^{l} y_i y_j a_i a_j k_{ij}$$

$$s.t. \quad \begin{cases} 0 \le a_i \le \dfrac{C}{l} \\ \sum_{i=1}^{l} y_i a_i = 0 \\ \sum_{i=1}^{l} a_i \ge Cv \end{cases} \qquad (2)$$

When $0 < a_i < \frac{C}{l}$, it is the normal support vector. Here $a_i = \frac{C}{l}$ is the bound support vector, and corresponding to $\xi_i > 0$ it stands for mistakenly dividing samples. Solve the formula (2) and optimum classification function (3) can be obtained.

$$f(x) = sign\{\sum_{i=1}^{l} a_i\, y_i k(x_i, x_j) + b\} \qquad (3)$$

Here $k(x_i, x_j) \equiv \varphi(x_i)^T \varphi(x_j)$ is the core function that is mainly divided into three types currently: polynomial core function $\varphi(x, x_i) = [(x, x_i) + 1]^d$, radial basis function $\varphi(x, x_i) = \exp\{-\frac{|x - x_i|^2}{\sigma^2}\}$ and sigmoid core function $\varphi(x, x_i) = \tanh(g(x, x_i) + p)$. And d is the order of the polynomial, σ is bigger than 0, p is the constant and b can be evaluated through the formula below:

$$b = \frac{1}{N_{sv}} \sum_{X_j \in J_N} \left(y_j - \sum_{X_i \in J_N} a_i y_i k(x_i, x_j) \right) \qquad (4)$$

Here J_N is the set of normal support vectors (NSV), J is the set of support vectors and N_{sv} is the number of support vectors.

To the bound support vectors, exist $a_i = \frac{C}{l}$ and suppose its number is N_{BSV}. From the constrain, if $\sum_{i=1}^{l} a_i = Cv$, we have $N_{sv} \times \frac{C}{l} \le \sum_{i=1}^{l} a_i$,

Namely $\frac{N_{BSV}}{l} \le v \qquad (5)$

and if $0 \le a_i \le \frac{C}{l}$, we have: $N_{SV} \times \frac{C}{l} \ge \sum_{i=1}^{l} a_i = Cv$,

Namely $v \le \frac{N_{SV}}{l} \qquad (6)$

From formula (5) and (6), we obtain

$$\frac{N_{BSV}}{l} \le v \le \frac{N_{SV}}{l} \qquad (7)$$

Seeing from the formula above, v is the lower bound of the ratio between the number of support vectors and the one of samples and also the upper bound of the ratio between the number of bound support vectors and the one of samples, so if v is determined, the accuracy of SVM (generalization ability) can be determined.

To the two classes of SVM, positive and negative, N_{BSV+} denotes the number of bound SVM belonging to the positive, N_{SV+} does the number of SVM belonging to the positive, N_{BSV-} does he number of bound SVM belonging to the negative, and N_{SV-} does the number of SVM belonging to the negative corresponding.

From formula (2), we have

$$\sum_{y_i=+1} a_i = \sum_{y_i=-1} a_i \qquad (8)$$

According to the formula above and formula (2), we have

$$\sum_{y_i=+1} a_i = \sum_{y_i=-1} a_i = \frac{1}{2} Cv \qquad (9)$$

Based on the formula above, there is $N_{SV+} \times \frac{C}{l} \ge \sum_{y_i=+1} a_i = \frac{1}{2} Cv$. To all of the support vectors, Namely $\frac{2N_{SV+}}{l} \ge v \qquad (10)$

and there is $N_{BSV+} \times \frac{C}{l} \le \sum_{y_i=+1} a_i = \frac{1}{2} Cv$. To all of the bound support vectors. Namely $v \ge \frac{2N_{BSV+}}{l} \qquad (11)$

From formula (10) and (11), we have

$$\frac{2N_{BSV+}}{l} \le v \le \frac{2N_{SV+}}{l} \qquad (12)$$

At the same time, to the negative class we can have

$$\frac{2N_{BSV-}}{l} \le v \le \frac{2N_{SV-}}{l} \qquad (13)$$

Suppose the numbers of positive and negative class of samples are l_+ and l_- respectively and $l = l_- + l_+$.

To formula (12) and (13), they can be denoted respectively:

$$\frac{N_{BSV+}}{l_+} \le \frac{v}{2}(1 + \frac{l_-}{l_+}) \le \frac{N_{SV+}}{l_-} \qquad (14)$$

and $\frac{N_{BSV-}}{l_-} \le \frac{v}{2}(1 + \frac{l_+}{l_-}) \le \frac{N_{SV-}}{l_-} \qquad (15)$

Seeing from formula (14) and (15), the class of big number of samples has smaller ratio of error classification and vice versa. When the difference between the numbers of positive and negative class of samples is very large, the one between the numbers of their support vectors is also very large, so the classification accuracy of different classes varies largely. Thus the accuracy of different classes can be modified through the weighted method in order to avoid large accuracy difference because of the numbers of samples.

3. THE ALGORITHM OF WEIGHTED SVM

3.1. Two Class of Weighted SVM

The problem of accuracy difference caused by the number of samples can be solved through weighting.

$$\min \quad \frac{1}{2}\|w\|^2 + C\left(-v\rho + \frac{1}{l}(\sum_{i=1}^{l_+} s_+\xi_i + \sum_{i=l_++1}^{l_++l_-} s_-\xi_i \right)$$
$$s.t. \quad \begin{cases} y_i(w^T\phi(x_i)+b) \ge \rho - \xi_i \\ \xi_i \ge 0, i=1,\cdots,l \\ \rho \ge 0 \end{cases} \tag{16}$$

Here s_i denotes weighted value of each sample and $0 < s_i \le 1$. To the positive and negative class, exist $s_i=s_+$ and $s_i=s_-$ respectively.

Its pairing form is

$$\min \quad \frac{1}{2}\sum_{i,j=1}^{l} y_i y_j a_i a_j k_{ij}$$
$$s.t. \quad \begin{cases} 0 \le a_i \le \frac{C}{l}s_+ & i=1,\cdots,l_+ \\ 0 \le a_i \le \frac{C}{l}s_- & i=l_++1,\cdots,l_++l_- \\ \sum_{i=1}^{l} y_i a_i = 0 \\ \sum_{i=1}^{l} a_i = Cv \end{cases} \tag{17}$$

And the relationship between the number of SVM and samples can be obtained.

$$\frac{s_+N_{BSV-}}{l_-} \le \frac{v}{2}(1+\frac{l_+}{l_-}) \le \frac{s_+N_{SV-}}{l_-}$$
$$\frac{s_-N_{BSV+}}{l_+} \le \frac{v}{2}(1+\frac{l_-}{l_+}) \le \frac{s_-N_{SV+}}{l_+}$$

In order to have the same accuracy, set

$$\frac{v}{2}(1+\frac{l_+}{l_-}) = \frac{v}{2}(1+\frac{l_-}{l_+})$$

and obtain
$$\frac{l_+}{l_-} = \frac{s_-}{s_+} \tag{18}$$

In the classification method of weighted SVM, classification accuracy can be improved through weighting small class of samples. Weighted SVM makes use of available information and does not increase the spending of calculating complication in the case of time and memory.

The key problem of weighted SVM is how to determine the value of $C_i(i=1, 2)$. C is too small, that is to say, the punishment of error classification is also too small and too many samples of error classification will come into being. Comparatively, C is too big, that is to say, the punishment of error classification is also too big and it can not play the compromise part.

Suppose there are two sample sets Ξ_1 and Ξ_2, whose sample numbers are N_1 and N_2 respectively and exist $N_1>N_2$. Rational value of C_i can be obtained through incessant approach, during which the formula below should come into existence:

$$\frac{1}{2}(w \cdot w) = \sum_{i=1}^{l} C_i\xi_i \tag{19}$$

From formula (18), the punishment of error classification to different class of samples is the inverse ratio to the number of samples. The initial value of C_i $(i=1, 2)$ can be obtained through the first scheme. From the discussion above, the new algorithm of SVM can be concluded as follows:

(1) To the side with more samples, afresh sampling according to equal probability and set the number $N_1=N_2$. (2) Choose bigger coefficient C and design SVM to the sample set$\{N_1, N_2\}$, then obtain the values of W and ξ_i. (3) According to formula (18) and (19), calculate C_1 and C_2.(4) Under the condition of new values of C_1 and C_2, redesign SVM to the sample set$\{N_1, N_2\}$ and obtain the values of W and ξ_i again. (5) Calculate current C_1 and C_2 according to the formula (10) and (11). (6) Judge the changes of C_1 and C_2.If the change is less than the initial threshold, turn to the step (7), or the step (4). (7) Obtain the ultimate SVM model.

Generally after three-time repetition the ultimate SVM model can be obtained.

3.2. Multi-Class Classifier Algorithm Based on SVM

As we know SVM can classify two classes efficiently, however in fact, we often have to classify more than two classes. Therefore we must generalize SVM to adapt to solve multi-class. There exists certain relationship between multi-class problem and two-class problem. If N classes can be classified, then any two classes among the N classes can be classified. So N classes can be surely classified by a certain number of two-class classifier.

Since SVM is excellent in separating two classes, we can combine SVM with 2-fork tree to construct the SVM decision tree to separate N classes [7]. The multi-class classifier based on decision tree can be efficiently divided into many two-class classifiers, while error is reduced. With the increase of classes, the number of classifiers increases linearly. We can efficiently get hierarchy structure of multi-class classifier regardless of prior knowledge. The calculation of classifier decreases along the increase of level.

At first SVM at the root is trained to classify samples into class 1 and the others, then SVM at next layer is trained to classify other class, and dividing them into class 2 and remaining. By analogy, remaining classes are separated by training remaining SVM in the end.

4. CLASSIFYING EXPERIMENTS AND DISCUSSION OF DEFECTS IN WELD IMAGE

Generally weld defects can be classified as crack, lack of penetration, lack of fusion, strip-shaped slag inclusion, spherical slag inclusion and pore 6 classes. The classifier based on the theory of weighted SVM is a general binary classifier which is not suitable for the cases of multi-classifier. In order to be able to solve the classifying problems of weld defects, each class of defects can train one SVM. In this way 6 weighted SVM obtained through training can classify weld defects as 6 classes, not just 2 classes.

In this paper the 84 welding pictures are selected as the total samples (totally 84×8 parameters), and the input vector $x_i = (x_{i1}, x_{i2}, \cdots x_{i8})(i = 1, 2, \cdots, 84)$, is composed of 8 feature parameters that must be normalized beforehand. The defect samples consist of 28 pore samples, 16 spherical slag inclusion samples, 10 strip-shaped slag inclusion samples, 10 lack of penetration samples, 10 lack of fusion samples, 10 crack samples. To 84 samples, 60% of them are used to train and 40% of them are use to test. Table1 shows the training and testing data of these samples.

In the experiments using weighted SVM to classify weld defects, Pore and the other 5 classes of weld defects have one mistake respectively and recognition accuracy arrives at 90.9％. Comparatively, using SVM to classify weld defects, crack has one mistake, lack of penetration has two mistakes, lack of fusion has one mistake and strip-shaped slag inclusion has two mistakes and the recognition accuracy arrives at 81.8%. Seeing from the experimental data, weighted SVM has higher recognition accuracy of weld defects than SVM does with less training samples, which accords to the analysis mentioned above. This indicates that weighted SVM can improve the classification accuracy of less samples effectively.

5. CONCLUSION

In this paper, through the analysis of the reason that the class difference does effects on the inclination in SVM algorithm, SVM can not obtain nice classifying ability when the numbers of samples vary greatly. Thus this paper puts forward a weighted SVM algorithm that applies for defect recognition in weld image. It can eliminate the effect caused by the class difference of training samples and decrease the errors of training samples, especially improve the classifying accuracy of small class of samples. Experimental results show that this algorithm can be used in the case of defect recognition in weld image effectively.

Table 1 6 Classes of Defect Samples

class	pore	spherical slag inclusion	strip-shaped slag inclusion	lack of penetration	lack of fusion	crack
Training data	17	10	6	6	6	6
Testing date	11	6	4	4	4	4

Reference

[1] T Warren, JiaWei Ni, "An automated radiographic NDT system for weld inspection-Flaw detection", NDT & E international, Vol 31, No.3, pp.183-192, 1998.

[2] Zhou wei, Wang chengxun, "Research and application of automatic recognition system to weld defects", Transactions of the China Welding Institution, Vol 13, No.1, pp.45-50, 1992.

[3] R R da Silva,M H S Siqueira,L P Caloba etc, "Radiographics Pattern Recognition of Welding Defects Using Linear Classifiers", Insight, Vol 43, No.10, pp.669-674, 2001.

[4] Vapnik V. The nature of statistical learning theory, New York, SPRINGER press, 1995.

[5] Scholkopf B, Smola A, Williamson R C et al. "New support vector algorithms", Neural Computation, Vol 12,No.5, pp. 1207~1245, 2000.

[6] Chang Chih-Chung, Lin Chih-Jen. "Training nu-support vector classifiers: theory and algorithms", Neural Computation, Vol 13, No.9, pp. 2119~2147,2001.

[7] Bennetr K P. Decision tree construction via linear programming (The Midwest Artificial Intelligence and Cognitive Science Society Conference), Utica, 1992.