

Research Article

Eliminating Vertical Stripe Defects on Silicon Steel Surface by $L_{1/2}$ Regularization

Wenfeng Jing, Deyu Meng, Chen Qiao, and Zhiming Peng

Institute for Information and System Science, Faculty of Science, Xi'an Jiaotong University, Xi'an 710049, China

Correspondence should be addressed to Chen Qiao, qiaochen@mail.xjtu.edu.cn

Received 2 July 2011; Accepted 19 October 2011

Academic Editor: Kwok W. Wong

Copyright © 2011 Wenfeng Jing et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The vertical stripe defects on silicon steel surface seriously affect the appearance and electromagnetic properties of silicon steel products. Eliminating such defects is a difficult and urgent technical problem. This paper investigates the relationship between the defects and their influence factors by classification methods. However, when the common classification methods are used in the problem, we cannot obtain a classifier with high accuracy. By analysis of the data set, we find that it is imbalanced and inconsistent. Because the common classification methods are based on accuracy-maximization criterion, they are not applicable to imbalanced and inconsistent data set. Thus, we propose a support-degree-maximization criterion and a novel cost-sensitive loss function and also establish an improved $L_{1/2}$ regularization approach for solution of the problem. Moreover, by employing reweighted iteration gradient boosting algorithm, we obtain a linear classifier with a high support degree. Through analyzing the classifier, we formulate a rule under which the silicon steel vertical stripe defects do not occur in the existing production environment. By applying the proposed rule to 50TW600 silicon steel production, the vertical stripe defects of the silicon steel products have been greatly decreased.

1. Introduction

Under normal process of silicon steel production, the surface of silicon steel products is smooth, as shown in Figure 1(a), but when production process is controlled imperfectly, the vertical stripes appear on the surface of the silicon steels at times, as shown in Figure 1(b). Such defects (briefly denoted as vertical stripe defects, or VSD in the following) not only affect the appearance effect of silicon steel, but also much degraded the lamination performance, resistance between layers and electromagnetic properties of silicon steel. How to eliminate the VSD problem has become one of the most important technical problems in silicon steel production.

In [1], the intrinsic mechanism of forming VSD was interpreted as follows: the high contents of Si and Al in silicon steel essentially lead to thick columnar crystals in the



Figure 1: (a) A normal silicon steel sheet; (b) a silicon steel sheet with vertical stripes.

casting slab organization, and thus α - γ phase transitions cannot occur in hot-rolling working procedure. Due to slow dynamic recovery and difficult recrystallization in later cold rolling and annealing process, such thick columnar crystals are hard to be completely broken. So, the vertical stripes arise on the surface of silicon steel products.

To eliminate the VSD, chemical analysis method and the equipment method have been generally employed in current silicon steel production [1]. The former method aims at enhancing the occurrence of α - γ phase transitions and accelerating the recrystallizing by empirically reducing the contents of Si and Al in silicon steel. The latter method eliminates the VSD mainly through the following ways: (i) adding a preannealing treatment; (ii) installing an electromagnetic stirring apparatus to enlarge the proportion of equiaxial crystals in casting plate slab; (iii) adding a normalization device in acidifying equipment group to increase the proportion of recrystallization of hot-rolled rolled sheet.

However, the above two methods suffer from some difficulties in the production of silicon steel [1]. The chemical analysis method cuts down the electromagnetic properties of silicon steel while reducing the contents of Si and Al. The equipment method needs a long period of technological reformation and huge investment. These difficulties above have hampered the further application of these methods.

In the process of silicon steel production, an amount of data of the VSD problem have been accumulated. We propose a data modeling method to search for the relationship between the VSD and its influence factors. Specifically, we transform the VSD problem into a special binary classification problem. For such a special classification problem, we cannot obtain an ideal classifier by common classification methods, such as the support vector machine [2], logistic regression [3], neural networks [4], and Fisher discriminant analysis [5], which are based on accuracy-maximization criterion.

In this paper, we propose a support-degree-maximization criterion instead of the common accuracy-maximization criterion and formulate a cost-sensitive loss function in place of usual loss function, and also we establish an $L_{1/2}$ regularization model to distinguish the key factors of the VSD. Furthermore, by utilizing the reweighted iteration gradient boosting algorithm, we obtain a linear model. In the model, the coefficients of some influence factors are very small or zeros. These results indicate that these factors are some less important influence factors and can be ignored under the existing production conditions. Through analyzing the coefficients of influence factors in the linear model, we put forward

a rule to avoid the VSD problem under existing production environment. Based on this rule, we propose an effective quality control strategy in the production of 50TW600 silicon steel.

The rest of this paper is arranged as follows: the VSD problem is formulated in Section 2. The methodologies, including the mathematical model and the related algorithm, are presented in Section 3. The experiments and their results are reported in Section 4. Finally, we conclude the paper with some useful remarks.

2. Problem Formulation

Silicon steel production is a very complicated process, including a series of working procedures, such as steel making, rough rolling, fine rolling, acid cleaning, rolling, annealing, coating, and cutting. In these working procedures, a large amount of complex chemical changes and physical changes occur continuously. As mentioned in [1], many working procedures are relevant to the VSD. The VSD problem is too difficult to model by the chemical and physical mechanism. So, we accumulated an amount of production data of the VSD to model the relationship between the VSD and its influence factors. But, there are too many influence factors related to the VSD, and the data of some influence factors are not easy to acquire. Therefore, by analyzing the arising mechanism of the VSD in 50TW600 silicon steel production [1], we select the 15 main factors and ignore some factors with little influence on the VSD, as listed in the following:

- (i) $T_1, T_2,$ and $T_3(^{\circ}\text{C})$: three tundish temperatures;
- (ii) $V_1, V_2,$ and $V_3 (\text{m} \cdot \text{min}^{-1})$: three casting speeds;
- (iii) $C_C, C_{\text{Si}}, C_{\text{Mn}}, C_{\text{S}}, C_{\text{P}},$ and $C_{\text{Al}}(\%)$: the contents of C, Si, Mn, S, P, and Al in silicon steel;
- (iv) $RT_0, FT_6(^{\circ}\text{C})$: rough rolling and fine rolling temperatures;
- (v) $CT(^{\circ}\text{C})$: coiling temperature.

According to the seriousness of the VSD, the products on production line of silicon steel are divided into two categories: (i) the negative class (majority class), labeled as “-1,” and the products in the class have no vertical stripe or slight vertical stripes which do not affect the physical properties of silicon steel. (ii) the positive class (minority class), labeled as “+1,” and the products in the class have serious vertical stripes which affect the physical performance of silicon steel. The negative class is acceptable to customers, while the positive class is unacceptable.

In order to eliminate the VSD, it is very necessary to make sure which factors have main influence on the VSD under existing control conditions of silicon steel production, and which factors are positive or negative influence. So, we need to model the relationship between the category of the VSD $y \in \{-1, +1\}$ and its influence factors $\mathbf{x} = (T_1, T_2, T_3, V_1, V_2, V_3, C_C, C_{\text{Si}}, C_{\text{Mn}}, C_{\text{S}}, C_{\text{P}}, C_{\text{Al}}, RT_0, FT_6, CT)^T$. This is a special binary classification problem because the problem has the following characteristics.

- (1) The data set of the VSD problem is imbalanced. In silicon steel products, qualified products (products without the VSD) are much more than unqualified ones (the products with the VSD).
- (2) The data set of the VSD problem is inconsistent. In the complicated process of silicon steel production, there are many factors intrinsically influencing the VSD.

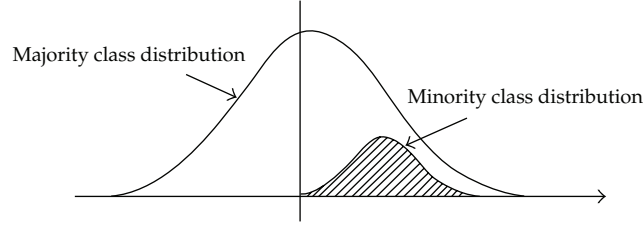


Figure 2: Class distributions of the imbalanced and inconsistent data set.

Through the analysis of the influence factors, we ignore lots of little influence factors and only select 15 main ones. The little influence factors ignored results in that there exist lots of samples whose numeric values of the influence factors are extremely close while their corresponding categories of the VSD are different.

In summary, the data set of the VSD problem is imbalanced and inconsistent, and its class distributions are shown in Figure 2.

In recent years, the knowledge discovery methods for inconsistent data concentrate on rough set approaches [6], and these methods are only suitable for discrete data. For continuous data, several methods for imbalance learning problems were reported [7]. These methods can be classified into two groups: resampling methods and algorithmic methods. Resampling methods rebalance class distributions by resampling the data space. And algorithmic methods strengthen to learn from minority class to improve common classification learning algorithms [2]. Further, extensive researches suggested that cost-sensitive methods in algorithmic methods are more superior than resampling methods in many application domains [4, 7–9]. But all methods mentioned above are not very effective to the VSD problem since the data set of the VSD problem is not only imbalanced but also inconsistent. In addition, the VSD problem needs to carry out variable selection. Recently, $L_{1/2}$ regularization methods were proposed and proved to be an effective classification approaches embedded variable selection [10]. Therefore, this paper investigates improved $L_{1/2}$ regularization methods by a cost-sensitive loss function to deal with the imbalanced and inconsistent data set of the VSD problem. The detail of the approach will be introduced in the following section.

3. Methodology

Before presenting our approach for the VSD problem, we need to introduce some preliminaries of the regularization methodology firstly.

3.1. The Regularization Framework of Classification Problems

Given a data set $S = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$, the binary classification problem can be modeled as follows [10]:

$$\hat{\beta}(\lambda) = \arg \min_{\beta} (R_{\text{emp}}(\beta) + \lambda P(\beta)), \quad (3.1)$$

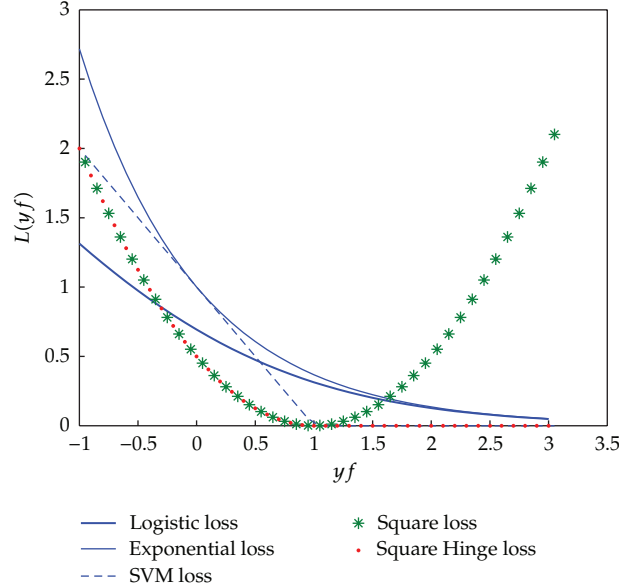


Figure 3: Comparison of common loss functions.

where $R_{\text{emp}} = (1/N) \sum_{i=1}^N L(y_i f(\mathbf{x}_i, \beta))$, $f(\mathbf{x}_i, \beta)$ is a given function with unknown parameter vector β , $L(yf)$ is a loss function about yf , $P(\beta)$ is a penalty term, and $\lambda (\geq 0)$ is a parameter tuning trade-off between R_{emp} and the penalty term $P(\beta)$.

Obviously, in (3.1), the following three elements are very important:

(a) the classification discriminant function $f(\mathbf{x}, \beta)$. Generally, given a set of basis functions $\mathcal{H} = \{h_1(\mathbf{x}), h_2(\mathbf{x}), \dots, h_p(\mathbf{x})\}$, where p is the number of the basis functions,

$$f(\mathbf{x}, \beta) = \sum_{j=1}^p \beta_j h_j(\mathbf{x}), \quad \beta_j \in \mathbb{R}^p. \quad (3.2)$$

In many applications, f is usually taken as the linear form, that is, $h_j(\mathbf{x}) = x_j$, $j = 1, 2, \dots, p$;

(b) the penalty term $P(\beta)$. When f is taken as a the linear function, the penalty function $P(\beta)$ is often formulated as

$$P(\beta) = \|\beta\|_k^k, \quad (3.3)$$

where $\|\beta\|_k$ is the k -norm of the coefficient vector of linear model;

(c) the loss function $L(yf)$. In binary classification problem, yf is defined as the margin. Usually, a loss function $L(yf)$ is a nonnegative and convex function about yf , such as logistic loss function $L(yf) = \log(1 + e^{-yf})$, exponential loss function $L(yf) = e^{-yf}$, SVM Hinge loss function $L(yf) = 1 + yf$ if $yf < 1$; 0 if $yf \geq 1$, square loss function $L(yf) = (1 - yf)^2$, and square Hinge loss function $L(yf) = (1 - yf)^2$ if $yf < 1$; 0 if $yf \geq 1$. The comparison of loss functions is shown in Figure 3.

From (3.1) and Figure 3, we can see that the different loss functions and different penalties result in different regularization algorithms. For example, in the Lasso [11],

the discriminant function f , the loss function, and the penalty term are taken as a linear function, the square loss function, and 1-norm of the coefficient vector of linear model, respectively. In the SVM [12], the three elements are taken as a linear function, square Hinge loss, and the square of 2-norm of the coefficient vector of linear model, respectively.

Based on the three important elements (a), (b), and (c) above, we will establish an $L_{1/2}$ regularization form to solve the VSD problem in the following.

3.2. $L_{1/2}$ Regularization Form for VSD Problem

In this subsection, we propose a support-degree-maximization criterion for an imbalanced and inconsistent data set and then give a cost-sensitive loss function to achieve support-degree maximization. And we preset linear function as classification discriminant function and take $\|\beta\|_{1/2}^{1/2}$ as the penalty term in regularization form the VSD problem.

(a) Support-degree-maximization criterion: in Section 2, we have pointed out that the data set of the VSD problem is imbalanced and inconsistent such that the problem cannot be solved by common classification methods. This is due to the fact that these methods employ the accuracy-maximization criterion. Hence, we propose a support-degree-maximization criterion. The notations used in this section are firstly given as follows.

In binary classification, the numbers of samples in positive class and negative class are denoted by N_+ and N_- , respectively, and the number of all samples is N , $N = N_+ + N_-$. After classifying, all samples are divided into four categories by the classifier, and the numbers of them are represented by TP, FP, TN, and FN [13].

In classification problem, accuracy is the most commonly used measure for assessing the capability of a classifier. It is defined as the ratio of the size of correctly classified samples to the size of the overall samples, that is,

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{N}. \quad (3.4)$$

It is clear that the common classification methods only expect maximizing accuracy of classifier. However, for the classification problem of the VSD, we cannot obtain enough high Accuracy by accuracy-maximization criterion since there exist too many inconsistent samples in the data set of the VSD problem.

So, we propose a support-degree-maximization criterion to obtain a classifier which can separate out a "good region," which contains as many negative samples as possible and almost no positive samples. In practical production, we can control production process parameters into the "good region" to eliminate the VSD.

Therefore, we define the Support degree and the Confidence degree as follows:

$$\begin{aligned} \text{Support} &= \frac{\text{TN}}{N}, \\ \text{Confidence} &= \frac{\text{TN}}{\text{TN} + \text{FN}}. \end{aligned} \quad (3.5)$$

Support represents the ratio of the size of correctly classified negative samples to the size of the overall samples, and Confidence measures the ratio of the size of correctly classified negative samples to the size of the overall classified as negative sample. In practical applications,

Confidence and Support are preset as a value near 100% and an acceptable value (e.g., 45%), respectively.

(b) Cost-sensitive loss function: to maximize Support, using a common loss function, we can construct a cost-sensitive loss function which gives a small penalty ρ to the false negative samples and a large penalty $1 - \rho$ to the false positive samples, $0 \leq \rho \leq 0.5$. Assume that $L(yf)$ is a loss function such as logistic loss, exponential loss, SVM Hinge loss, and square Hinge loss, then the cost-sensitive loss function can be constructed as follows:

$$\hat{L}(y_i f_i, \rho) = \begin{cases} \rho L(y_i f_i) & \text{if } y_i = -1, \\ (1 - \rho) L(y_i f_i) & \text{if } y_i = +1. \end{cases} \quad (3.6)$$

For simplicity, expression (3.6) can also be written as

$$\hat{L}(y_i f_i, \rho) = \frac{1}{2} ((1 + y_i)(1 - \rho) + (1 - y_i)\rho) L(y_i f_i). \quad (3.7)$$

(c) Linear classification discriminant function: in order to facilitate distinguishing the factors with little influence on the VSD, the linear function is employed as the classification discriminant function

$$f(\mathbf{x}, \beta) = \sum_{i=1}^p \beta_i x_i, \quad (3.8)$$

where p is the dimensionality of \mathbf{x} and β_i is the parameter to be determined. The sign of β_i represents that x_i positively or negatively influences the VSD, and $|\beta_i|$ represents the influence extent of x_i to the VSD, $i = 1, 2, \dots, p$.

(d) Penalty term: in [10], Xu et al. pointed out that $L_{1/2}$ regularizer is a good representative of L_p ($0 < p < 1$). Because when $1/2 \leq p < 1$, the $L_{1/2}$ regularizer always yields the best sparse solution, and when $0 < p < 1/2$, the sparse property of L_p regularizer is similar to that of the $L_{1/2}$ regularizer. Therefore, in order to determine the variables with greater influence on the VSD under existing production control state, the penalty term in regularization frame is adopted as follows:

$$\|\beta\|_{1/2}^{1/2} = \sum_{i=1}^p |\beta_i|^{1/2}. \quad (3.9)$$

Based on the above discussion (a)~(d), the regularization model for the VSD problem can be formulated as

$$F_S = \arg \min_{\beta \in \mathbb{R}^p} \left(\frac{1}{N} \sum_{i=1}^N \hat{L}(y_i f_i, \rho) + \lambda \|\beta\|_{1/2}^{1/2} \right), \quad (3.10)$$

where $\lambda \geq 0$ is the regularization parameter, \hat{L} , f and $\|\beta\|_{1/2}^{1/2}$ are defined as formulas (3.7), (3.8), and (3.9), respectively.

By integrating the reweighted iteration strategy, an effective and efficient algorithm for the optimization problem (3.10) can be designed, as introduced in the following subsection.

3.3. Reweighted Iteration Algorithm for the Proposed Model

As mentioned in [10], $L_{1/2}$ regularizer can yield more sparse solution than L_1 regularizer. Nevertheless, $L_{1/2}$ regularization is more difficult to be solved than L_1 regularization because the former is a nonconvex optimization problem, while the later is convex optimization problem. For the $L_{1/2}$ regularization problem (3.10), based on reweighted iteration, algorithm we propose an effective and efficient algorithm. Its main idea is to transform an $L_{1/2}$ regularization problem into a series of L_1 regularization problems which can be solved effectively by existing L_1 regularization algorithms, like gradient boosting algorithm. The algorithm is described as follows.

Algorithm 1 (reweighted iteration algorithm for the VSD problem).

Step 1. Initialize $\beta^{(0)} = (1, 1, \dots, 1)^T$, and set the maximum iteration step K . Set iteration step $t = 0$, Support = S_0 , and Confidence = C_0 .

Step 2. Apply L_1 gradient boosting algorithm to solve

$$\beta^{(t+1)} = \arg \min \left\{ \frac{1}{N} \sum_{i=1}^N \hat{L}(y_i f_i, \rho) + \lambda \sum_{i=1}^p \frac{1}{\sqrt{|\beta_i^{(t)}|}} |\beta_i| \right\}, \quad (3.11)$$

and set $t = t + 1$.

Step 3. Compute Support^(t) and Confidence^(t) if $t > K$ or (Support^(t) $\geq S_0$ and Confidence^(t) $\geq C_0$) output $\beta^{(t)}$. Otherwise, go to Step 2.

In the above algorithm, the initial value $\beta^{(0)}$ is taken as $(1, 1, \dots, 1)^T$, and thus in the first iteration ($t = 0$), Step 2 exactly solves an L_1 regularization problem. When $t \geq 1$, Step 2 needs to solve a reweighted L_1 regularization problem, which can be transformed into an L_1 regularization via linear transformation. It should be noted that some coefficients of $\beta^{(t)}$ are zeros when $t \geq 1$. In order to guarantee the feasibility, we replace $1/\sqrt{|\beta_i^{(t)}|}$ with $1/\sqrt{|\beta_i^{(t)}| + \varepsilon_0}$ in Step 2, where ε_0 is any fixed positive real number. In addition, S_0 and C_0 can be set as expected values, for example, $S_0 = 45\%$, $C_0 = 100\%$. In this algorithm, L_1 regularization problem is solved by ε -gradient boosting algorithm (see detail in the appendix).

4. Solution of the VSD Problem and Its Application

In this section, we carry out numerical experiments on the VSD problem (3.10) using reweighted iteration algorithm.

4.1. Experiments and Results

(a) Data preparation: we collected the samples from 50TW600 silicon steel products for 3 months when the rate of VSD products was as high as 12.1%. We used the samples of the first two months as training set and the those of the third month as testing set. After discarding the samples with null values, the training set has 3303 samples, including 3195 negative class samples and 108 positive class ones. The proportion of the negative to the positive class samples is 29.58:1. Obviously, the class distribution of the data set is greatly imbalanced. Moreover, the 108 positive class samples are almost all inconsistent samples. The testing set has 1026 samples, including 981 negative class samples and 45 positive class ones.

(b) Data standardization: to avoid the impact of numerical scale on computational precision, the data of every independent variable is standardized as mean and standard deviation being 0 and 1, respectively.

(c) Experimental result: we used regularization formation (3.10) and Algorithm 1 for the standardized data set. In this experiment, we preset $S_0 = 45\%$, $C_0 = 100\%$, $K = 20$, and $\varepsilon_0 = 0.000001$. In order to maximize support degree, ρ is preset as a very small value 0.00001. And we employed logistic loss, SVM Hinge loss, exponential loss, and square Hinge loss function in cost-sensitive loss function, respectively. We obtained the best linear classifier with the maximal support degree 47.74% when square Hinge loss function was employed.

The obtained linear classifier is

$$f(\mathbf{x}') = \text{sign}(\beta' \cdot \mathbf{x}'), \quad (4.1)$$

where $\beta' = [3.848, 0.0, 1.422, 2.527, 4.482, -1.532, -1.168, 6.693, -1.417, 4.459, -1.704, 4.757, -0.766, -6.290, -0.640]^T$, $\mathbf{x}' = [T'_1, T'_2, T'_3, V'_1, V'_2, V'_3, C'_C, C'_{Si}, C'_{Mn}, C'_P, C'_S, C'_{Al}, RT'_0, FT'_6, CT']^T$. By inverse standardizing transform, classifier (4.1) obtained from the standardized data can be transformed into classifier (4.2) in original data

$$F(\mathbf{x}) = \text{sign}(\beta \cdot \mathbf{x} + b), \quad (4.2)$$

where $\beta = [0.575, 0.0, 0.205, 35.695, 212.011, -68.283, -2548.028, 213.512, -83.821, 1657.123, -1796.934, 234.172, -0.115, -0.896, -0.276]^T$, $\mathbf{x} = [T_1, T_2, T_3, V_1, V_2, V_3, C_C, C_{Si}, C_{Mn}, C_P, C_S, C_{Al}, RT_0, FT_6, CT]^T$, and $b = -601.536$. The effect of the classifier $F(\mathbf{x}) = \text{sign}(\beta \cdot \mathbf{x} + b)$ is shown in Figure 4. The horizontal axis is $\beta \cdot \mathbf{x} + b$, and the vertical axis is FT_6 . The purpose of using FT_6 as the vertical axis is to show the classification result on the two-dimensional plane.

The confusion matrix of the best linear classifier is shown in Table 1.

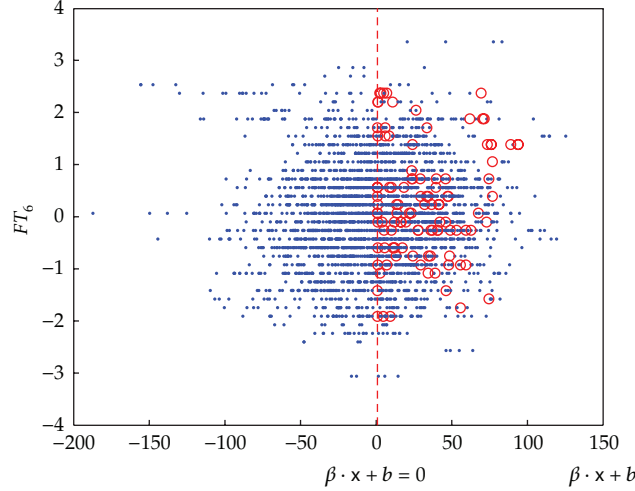
From Figure 4 and Table 1, it is found that the left of the classifier (the dotted line) is the optimal region of the silicon steel production. This is because there are no defective products and a great many quality products in the area.

(d) Result analysis: from the experimental results, we can draw the following conclusions.

- (1) Since the linear classifier (4.1) has high enough support degree (47.74%) and confidence degree (100%), the classifier can be put into practical application.
- (2) In linear classifier (4.2), the absolute values of $\hat{\beta}'_{13}$ and $\hat{\beta}'_{15}$ are very small and $\hat{\beta}'_2$ is 0. So the influences of the corresponding factors RT_0 , CT , and T_2 can be ignored under the existing production status.

Table 1: Confusion matrix.

Predicted positives	Predicted negatives
Real positives 108	0
Real negatives 1577	1726

**Figure 4:** Experimental results for solving the VSD problem.

- (3) From Figure 4, we see that $\beta \cdot x + b < 0$ can be regarded as a discrimination rule. If a group of production control parameters x satisfy the rule, then it is almost impossible for the VSD problem to arise. However, according to the practical experience and theoretical analysis, $\beta \cdot x + b$ cannot be too small. Otherwise, the electromagnetic properties of silicon steel will be reduced.
- (4) Evaluating model: using the testing data set, we verify the model (4.1). When 45 positive class samples are put in model (4.1), it is found that all $\beta \cdot x + b > 0$. In 981 negative class samples, 523 samples satisfy $\beta \cdot x + b < 0$ and 458 samples have $\beta \cdot x + b > 0$. That is to say, the model (4.1) separates out a “good region” with enough many quality products of silicon steel production.

4.2. Improved Control Strategy of Silicon Steel Production

During the period that the silicon steel vertical stripes arose frequently, the control strategy used in the production of silicon steel is shown in Table 2. The hitting target values are the expected values of corresponding process parameters. Unfortunately, the hitting target values satisfy $\beta \cdot x + b > 0$. Under such a poor product control strategy, the possibility of generating VSD problem was very high, and the rate of the silicon steel products with VSD problem reached a higher value 12.1%.

According to the result analysis of the experiments presented in Section 4.1, we suggest a production control strategy with improved hitting target values

$C_{Si} = 1.36, C_{Mn} = 0.40, C_P = 0.07, C_{Al} = 0.25$, and $FT_6 = 850$. In order to preserve the stability of the silicon steel productions, the hitting target values of the variables $T_1, T_2, T_3, V_1,$

Table 2: The production control strategy when the VSD problem was very serious.

Control parameters	T_1	T_2	T_3	V_1	V_2	V_3	C_C	C_{Si}
Original hitting target values	1540	1540	1540	0.75	0.75	0.75	0.003	1.38
Control parameters	C_{Mn}	C_P	C_S	C_{Al}	RT_0	FT_6	CT	
Original hitting target values	0.32	0.09	0.0035	0.27	960	830	655	

V_2 , V_3 , C_C , C_S , RT_0 , and CT are unchanged since they have little influence on VSD problem of silicon steel surface under the existing production conditions. It is easy to verify that the improved hitting target values satisfy the rule $\beta \cdot x + b < 0$.

When the improved production control strategy is applied to the production of 50TW600 silicon steel, the rate of the silicon steel products with vertical stripes has been lowered to a level less than 1.8%.

5. Conclusion

In this work, through analyzing the data set of the VSD problem, it is found that such data set is imbalanced and inconsistent, and common classification methods based on classification accuracy are not suitable for this specific classification problem. For this reason, a new classification criterion called support-degree-maximization criterion for the imbalanced and inconsistent data sets has been proposed. Moreover, to distinguish the factors with little influence on the VSD, $L_{1/2}$ regularization form has been established for the VSD problem. By solving the regularization problem with reweighted iteration gradient boosting algorithm, the rule of avoiding silicon steel vertical stripes under existing production environment has been put forward. Furthermore, an improved production control strategy has been suggested and applied to the silicon steel production line. As a result, the rate of the products with vertical strip defects is greatly alleviated. Although the VSD problem has decreased greatly, there still exist some defect products up to about 1.8%, which is caused by the rise and fall of control variable values. Therefore, it is necessary to enhance hitting rates of the target values of the influence factors by means of 6σ management approach. This is a work in our future research.

Appendix

Gradient Boosting Algorithm for L_1 Regularization

L_1 regularization problem is formulated as follows:

$$F_S = \arg \min_{f \in \mathcal{F}} \left(\frac{1}{N} \sum_{i=1}^N L(y_i f(x_i, \beta)) + \lambda \|\beta\|_1 \right), \quad (\text{A.1})$$

where $\lambda \geq 0$ is the regularization parameter, L is a loss function, $\|\beta\|_1 = \sum_{i=1}^p |\beta_i|$, and $\mathcal{F} = \{f : f(\mathbf{x}, \beta) = \sum_{i=1}^p \beta_i h_i(\mathbf{x}), \beta_i \in R^p\}$, where h_1, h_2, \dots, h_p are a group of basis functions.

Gradient boosting algorithm [14] is well known as one of the most effective algorithms for L_1 regularization problems. The algorithm begins with $\beta^{(0)} = (0, 0, \dots, 0)^T$, that is, $F_{(0)}(x) = 0$, and after T steps iterations, the following model can then be obtained:

$$F_T(\mathbf{x}) = \sum_{t=1}^T \alpha_t h_{j_t}(\mathbf{x}), \quad (\text{A.2})$$

where $\beta_j^{(T)} = \sum_{i=j}^T \alpha_i$. If we have the combination model $F_{t-1}(\mathbf{x})$ of the first $t-1$ steps, then we want to seek an $h_{j_t}(\mathbf{x})$ from $\{h_j(\mathbf{x})\}_1^p$ such that empirical risk $\sum_{i=1}^N L(\mathbf{y}_i F)$ decreases quickly. Denote $\beta_{j_t}^{(t)} = \beta_{j_t}^{(t-1)} + \alpha_t$, $\beta_k^{(t)} = \beta_k^{(t-1)}$, $k \neq j_t$, then $F_t(\mathbf{x}) = F_{t-1}(\mathbf{x}) + \alpha_t h_{j_t}(\mathbf{x})$ can be obtained by maximizing first-order descendent quantity of the loss function in the negative gradient direction, that is,

$$h_{j_t} = \arg \max_{h_j \in \mathcal{H}} \left| \sum_{i=1}^N \frac{\partial L(\mathbf{y}_i F)}{\partial \beta} \right|_{F=F_{t-1}}. \quad (\text{A.3})$$

Obviously, the expression above is equivalent to

$$h_{j_t} = \arg \max_{h_j \in \mathcal{H}} \left| \sum_{i=1}^N \left(\frac{\partial L(\mathbf{y}_i F)}{\partial F} h_j(\mathbf{x}_i) \right) \right|_{F=F_{t-1}}. \quad (\text{A.4})$$

Subsequently, we choose the direction $\text{sign}(\alpha_t) = \text{sign}(\sum_{i=1}^N -(\partial L(\mathbf{y}_i F)/\partial F) h_j(\mathbf{x}_i))|_{F=F_{t-1}}$ as descendent direction, and the descendent quantity along this direction $\gamma_t = |\alpha_t|$ can be obtained by solving the following optimization problem:

$$\gamma_t = \arg \min_{\gamma} \sum_{i=1}^N L(\mathbf{y}_i (F_{t-1} + \gamma h_{j_t}(\mathbf{x}_i))). \quad (\text{A.5})$$

Hastie et al. [15] pointed out that the above linear search method is too greedy that the algorithm would lose its stability. For this reason, a "learning slowly" skill was proposed by assigning a descendent step size $\gamma_t = \varepsilon$, $\varepsilon \in (0, 1)$, which is called ε -Boosting algorithm.

Summarizing the analysis above, the normal framework of gradient boosting algorithm for L_1 regularization can be described as follows.

Algorithm 2 (gradient boosting algorithm).

Step 1. Initialize the coefficients of basis functions.

Set iteration number $t = 0$, and the coefficients of basis functions $\beta_j^{(0)} = 0$, $j = 1, 2, \dots, p$, p is the number of basis functions, and descendent step size $\gamma_t = \varepsilon$.

Step 2. Calculate the current fitting value.

Set $t = t + 1$, and calculate the current fitting value $F_i^{(t)} = \sum_{j=1}^p \beta_j^{(t-1)} h_j(\mathbf{x}_i)$, $i = 1, \dots, N$.

Step 3. Select a basis function.

Calculate the negative gradient of the loss function at each sample $w_i = -\partial L(\mathbf{y}; F) / \partial F|_{F=F_i^{(t)}}$, $i = 1, \dots, N$, and determine $j_t = \arg \max_j |\sum_{i=1}^N w_i h_j(\mathbf{x}_i)|$, $j \in \{1, 2, \dots, p\}$.

Step 4. Adjust the coefficients of basis functions.

$$\text{Set } \beta_{j_t}^{(t)} = \beta_{j_t}^{(t-1)} + \gamma_t \text{sign}(\sum_{i=1}^N w_i h_{j_t}(\mathbf{x}_i)), \beta_k^{(t)} = \beta_k^{(t-1)}, k \neq j_t.$$

Step 5. If $t < T$, go to Step 2, otherwise output $\beta^{(t)}$.

Acknowledgments

The authors are grateful to their collaborators of Iron & Steel (Group) Co., Ltd., for their provision of the silicon steel data, and for their enthusiasm, encouragement, and profound viewpoints in this work. This research is supported by the China NSFC project under Contract 60905003.

References

- [1] Z. S. Xia, "Research and production of boron bearing cold rolled non-oriented silicon steel," *Shanghai Metals*, vol. 25, no. 4, pp. 20–23, 2003.
- [2] Y. Lin, Y. Lee, and G. Wahba, "Support vector machines for classification in nonstandard situations," *Machine Learning*, vol. 46, no. 1–3, pp. 191–202, 2002.
- [3] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression*, John Wiley & Sons, Hoboken, NJ, USA, 2000.
- [4] X. Y. Liu and Z. H. Zhou, "Training cost-sensitive neural networks with methods addressing the class imbalance problem," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 1, pp. 63–77, 2006.
- [5] G. Fung, M. Dundar, J. Bi et al., "A fast iterative algorithm for Fisher discriminant using heterogeneous kernels," in *Proceedings of the 21st international conference on Machine learning (ICML '04)*, Banff, AB, Canada, 2004.
- [6] W. X. Zhang, J. S. Mi, and W. Z. Wu, "Approaches to knowledge reductions in inconsistent systems," *International Journal of Intelligent Systems*, vol. 18, no. 9, pp. 989–1000, 2003.
- [7] H. B. He and A. Edwardo, "Learning from imbalanced data," *IEEE Transactions on Knowledge and Data Engineering In Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, 2009.
- [8] K. McCarthy, B. Zabar, and G. M. Weiss, "Does cost-sensitive learning beat sampling for classifying rare classes?" in *Proceedings of the 1st International Workshop on Utility-Based Data Mining*, pp. 69–77, 2005.
- [9] X. Y. Liu and Z. H. Zhou, "The influence of class imbalance on cost-sensitive learning: an empirical study," in *Proceedings of the 6th International Conference on Data Mining (ICDM '06)*, pp. 970–974, Hong Kong, 2006.
- [10] Z. B. Xu, H. Zhang, Y. Wang, X. Chang, and Y. Liang, " L_{12} regularization," *Science in China F: Information Sciences*, vol. 53, no. 6, pp. 1159–1169, 2010.
- [11] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *Journal of the Royal Statistical Society. Series B. Methodological*, vol. 58, no. 1, pp. 267–288, 1996.
- [12] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, NY, USA, 1995.
- [13] Y. Sun, *Cost-sensitive boosting for classification of imbalanced data*, Ph.D. thesis, University of Waterloo, Waterloo, ON, Canada, 2007.
- [14] S. Rosset, J. Zhu, and T. Hastie, "Boosting as a regularized path to a maximum margin classifier," *Journal of Machine Learning Research*, vol. 5, pp. 941–973, 2003/04.
- [15] T. Hastie, R. Tibshirani, and J. Friedman, "The elements of statistical learning: data mining," in *Inference and Prediction*, Springer-Verlag, New York, NY, USA, 2001.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

