Cost-sensitive Multi-class SVM with Reject Option: A Method for Steam Turbine Generator Fault Diagnosis

Chao Zou, En-hui Zheng, Hong-wei Xu, Le Chen

***Abstract*—The steam turbine generator faults not only damage the generator itself, but also cause outages and loss of profits. The traditional fault diagnosis systems care only about high diagnosis accuracy. But different misdiagnoses may lead to quite different losses and it is unreliable if misdiagnoses were accepted. In order to reduce the total loss caused by misdiagnoses and improve the diagnosis reliability, in this paper, cost integrated multiclass SVM with reject option (CIMCR-SVM) is proposed. Firstly, we present a very simple and effective method to make the multi-class classifiers cost-sensitive. Secondly, diagnosis reliabilities were evaluated by a reliability evaluator, and reject option is introduced for rejecting classified samples with lower diagnosis reliabilities. Experimental results demonstrate that CIMCR-SVM is able to minimize the average cost and improve the diagnosis reliability.**

***Index Terms*—SVM, multi-class, cost-sensitive, fault diagnosis, reject option.**

I. Introduction

Steam turbine generators are major components of the thermal power plants and nuclear power plants to convert heat energy of steam into electrical energy through mechanical energy. Vibration phenomena appear at the rotor and the bearing if something is wrong with the turbine generators [1, 2]. There are several types of faults such as rotor unbalance faults, centering faults, oil whipping faults, etc. The generator faults not only damage the generator itself, but also cause outages and loss of profits [2]. Moreover, misdiagnosing may also lead to serious losses.

Several approaches to the identification of operation anomalies have been investigated. In general, these approaches are based on techniques of pattern recognition which aim at inferring the different faulty and anomalous system states from the corresponding different patterns of evolution that the involved process variables follow. In particular, a large number of diagnostic systems of steam turbine have been proposed in the past, based on the advances of machine learning techniques such as fuzzy logic system [3], genetic algorithm [4], artificial neural network [4], and Support vector machines [5, 6]. These systems are effective on their diagnostic scope, and some of them are cost sensitive [1-6]. Unfortunately, these systems ignore inter-class relations with considering misdiagnosis cost. This is why some other researchers preferred to work with multiclass classifiers.

Support Vector Machines (SVM) were originally designed for binary classification, the most widely used implementation to extend binary classifiers for multi-class classification is construct a multiclass classifier by combining several binary classifiers [8], i.e., One against All method (OAA) and One against One method (OAO) [6-8]. For the purpose of minimizing the total cost, we introduce the cost-sensitive learning in this works. In the machine learning community, cost-sensitive learning has been studied for years [9-15]. There are two types of methods to make cost-sensitive classifiers. One is to design cost-sensitive learning algorithm directly. The other is to convert existing standard classifiers to cost-sensitive ones. The latter is also called meta-learning [10, 11] and has been used widely, i.e., sampling [10, 13], thresholding [10, 11], weighting [10, 12, 14] and so on. Because error-cost is class-dependent: different errors may lead to quite different losses, the ambiguous classifications will reduce the reliability of the classifiers. It is meaningful to reject low reliability decisions with a reject option [15-19]. The reject option involves two stages: 1) evaluating the classification reliability; 2) rejecting the unreliable classified samples with an optimum threshold [15] or a fixed one [16, 17].

In this paper, we propose a cost integrated multiclass support vector machine with reject option (CIMCR-SVM) which minimize the average cost and reject the unreliable classifications. Experimental results show that CIMCR-SVM is able to reduce the average diagnosis cost and improve the diagnosis reliability.

The rest of this paper is organized as follow: in Section II, standard SVM and multi-class SVM are reviewed. In Section III, we describe the method of cost integration, cost-sensitive multi-class SVM classifiers and reject option. The numerical experiments of diagnosis test are presented in Section IV. Finally, we conclude the paper in Section V.

Manuscript received December 18, 2009. This work was supported by the Natural Science Foundation of Zhejiang Province P. R. China under Grant Y1080950, the National Natural Science Foundation of P. R. China under Grant 60905034 and the National Department Public Benefit Research Foundation of P. R. China under Grant 2007GYJ016.

Chao Zou is with the China Jiliang University, Hangzhou, CO 310018 P. R. China (e-mail: zouc@cjlu.edu.cn).

En-hui Zheng, corresponding author of this paper, is with the China Jiliang University, Hangzhou, CO 310018 P. R. China (phone: +86-571-86914549 5; e-mail: ehzheng@cjlu.edu.cn).

Hong-wei Xu is with the China Jiliang University, Hangzhou, CO 310018 P. R. China (e-mail: xhw@cjlu.edu.cn).

Le Chen is with the China Jiliang University, Hangzhou, CO 310018 P. R. China (e-mail: cl7788@126.com).

IACSIT International Association of Computer Science and Information Technology WWW.IACSIT.ORG

II. Reviews of Support Vector Machines

Support Vector Machine (SVM) motivated by statistical learning theory, provides a novel approach to the two-category classification problem [7]. SVM has been successfully applied to a number of applications ranging from face identification, fault detection, medical diagnosis, text categorization, and so on.

A. Basic SVM (binary classifier)

The set of labeled training data

$$\begin{aligned}&(x_1, y_1), ..., (x_i, y_i), ..., (x_l, y_l),\\ &x_i \in \mathrm{R}^n,\ y_i \in \{-1, +1\},\ i = 1, ..., l,\end{aligned} \tag{1}$$

is said to be linearly separable if there exists a vector $\mathbf{w}$ and a scalar b such that the $y_i(\mathbf{w} \bullet \mathbf{x_i} + b) \geq 1$ is valid for all elements of the training set (1) [8].

The optimal hyperplane $\mathbf{w}_0 \cdot x + b_0 = 0$ is the unique one which separates the training data with a maximal margin [7]. The maximal distance between the projections of the training vectors of two different classes is given by maximizing

$$\rho(\mathbf{w}_0, b_0) = \frac{2}{\|\mathbf{w}_0\|}. \tag{2}$$

When the training data can not be separated without error, Cortes and Vapnik (1995) introduced the non-negative variable $\xi_i > 0$, $i = 1, ..., l$. The soft margin hyperplane can be found by minimizing

$$\begin{aligned}&\frac{1}{2}\|\mathbf{w}\|^2 + C \cdot \sum_{i=1}^{l} \xi_i,\\ &s.t.\ \ y_i(\mathbf{w} \bullet \mathbf{x_i} + b) \geq 1 - \xi_i,\ i = 1, ..., l,\end{aligned} \tag{3}$$

where constant $C \geq 0$ is the error penalty.

The minimized (3) determine the hyperplane that minimizes the number of errors on the training set and separate the rest of the elements with maximal margin.

SVM can also be used in non-linear classification tasks with application of *Kernel Function*s, such as linear, polynomial and Gaussian RBF [7]. The input data is mapped onto a high-dimensional feature space, where the linear classification is possible.

B. Multi-class SVM (MC-SVM)

SVM was originally designed for binary classification. Several methods have been proposed to effectively extend the approach to multiclass classification [6-8]. The most widely used implementation for multi-class classification are One against All method (OAA) and One against One method (OAO) [6-8]. OAA method constructs k SVM models where k is the number of classes. The i th SVM is trained with all of examples in the i th class with positive labels, and all the other examples with negative labels. OAO method constructs $k(k-1)/2$ classifiers where each one is trained on data from two classes. The decision is made using the voting strategy. The voting approach described above is also called as Max Win strategy [6].

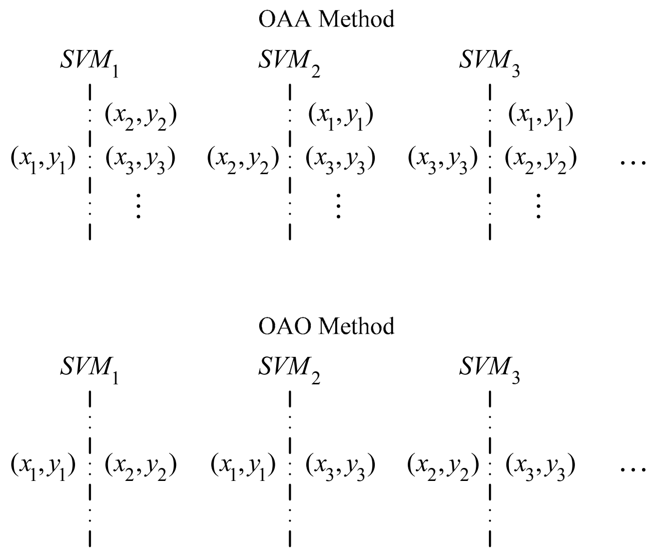

Figure 1. The diagram of the OAA and OAO method

III. Cost Integrated MC-SVM with Reject Option

A. Cost Integrated Multi-class Classifier Design

In fault diagnosis applications, the cost of misclassifying a fault instance is fairly more expensive than that of misclassifying a normal one [13]. When thinking in terms of misclassification costs, it is easy to posit a cost matrix that is logically contradictory because not all entries in the matrix are measured from the same baseline. Mathematically, let the (t, s) entry in a cost matrix C_{mis} be the cost of predicting class s when the true class is t. If $s = t$ then the prediction is correct and the misclassification cost is 0, while if $s \neq t$ the prediction is incorrect and the corresponding misclassification cost should be a positive number. The cost information is given by domain knowledge and appears as precise value.

In order to outweigh errors on classes with higher misclassification cost, we propose a very simple, yet general and effective method to improve the sensitivity of the OAA and OAO classifiers. That is, multiply the penalty term by misclassification cost. The specific method is described as follow.

1) Cost Integrated One against All method (CI-OAA)

Given l training data $(x_1, y_1), ..., (x_i, y_i), ..., (x_l, y_l)$, where $x_i \in \mathrm{R}^n$, $i = 1, ..., l$ and $y_i \in \{1, ..., k\}$ is the class label of x_i, the i th SVM can be obtained as a solution to the following optimization problem:

$$\min \frac{1}{2}\left\|\mathbf{w}^i\right\|^2 + C\sum_{i=1}^{l}\left(\frac{1}{k}\sum_{s=1}^{k}C_{mis}(t,s)\right)\cdot \xi_j^i,$$
$$s.t. \begin{cases}(\mathbf{w}^i)^T\cdot\phi(\mathbf{x}_j)+b^i \geq 1-\xi_j^i, & if \quad y=i,\\ (\mathbf{w}^i)^T\cdot\phi(\mathbf{x}_j)+b^i < -1+\xi_j^i, & if \quad y\neq i,\end{cases} \tag{4}$$
$$\xi_j^i \geq 0,\; i,j=1,\dots,l,\; s,t=1,\dots,k,$$

where ξ_j^i is the slack variable, C is the error penalty, $C_{mis}(t,s)$ is the misclassification cost matrix and the training data $\mathbf{x}_i$ is mapped onto a higher-dimensional feature space by function ϕ.

2) Cost Integrated One against One method (CI-OAO)

For training data form the i th and the j th classes, we solve the following problem:

$$\min \frac{1}{2}\left\|\mathbf{w}^{ij}\right\|^2 + C\sum_{t}C_{mis}(j,i)\cdot \xi_t^{ij},$$
$$s.t. \begin{cases}(\mathbf{w}^{ij})^T\cdot\phi(\mathbf{x}_t)+b^{ij} \geq 1-\xi_t^{ij}, & if \quad y_t=i,\\ (\mathbf{w}^{ij})^T\cdot\phi(\mathbf{x}_t)+b^{ij} < -1+\xi_t^{ij}, & if \quad y_t=j,\end{cases} \tag{5}$$
$$\xi_t^{ij} \geq 0,\; i,j=1,\dots,l,$$

where ξ_t^{ij} is the slack variable, C is the error penalty and $C_{mis}(j,i)$ is the misclassification cost (in which i is the predict class and j is the actual class).

B. Classification Reliability Evaluation

1) Post Probability

Both OAA and OAO exploit the fact that the absolute value of the output of a SVM is proportional to the distance of an input instance form the class boundary estimated by the classifier [15].

$$d(x_i \mid SVM_j) = \left(\mathbf{w}_i\cdot x_i + b_i\right)/\left\|\mathbf{w}_i\right\|. \tag{6}$$

Equation (6) denotes the distance between x_i and the boundary of the j th SVM_j.

The post probability of x_i belongs to the class j can be estimated by the *Sigmoid Function*:

$$P_j^i(x_i) = \frac{1}{1+\exp\left(-a\cdot d(x_i \mid SVM_j)\right)}, \tag{7}$$

where a is the parameter of *sigmoid function*.

2) Classification Reliability

Classification reliability can be expressed by associating two situations [16, 17]: 1) there is no class whose post probability is sufficient to judge the classification reliable; 2) two or more classes whose post probability are so similar that there is not a clear overwhelming class. The two situations mentioned above can be denoted by ψ_a and ψ_b respectively. For the purpose of providing a comprehensive measure of the reliability of a classification, the *Reliability Function* [16] is defined as follow:

$$\psi(x_i) = \frac{1}{2}\left(\psi_a(x_i)+\psi_b(x_i)\right),$$
$$\text{also can be: } \psi(x_i) = \min\left(\psi_a(x_i),\psi_b(x_i)\right), \tag{8}$$
$$\text{or: } \psi(x_i) = \max\left(\psi_a(x_i),\psi_b(x_i)\right),$$
$$\psi_a(x_i) = \pi_1(x_i),\; \psi_a(x_i) = 1-\pi_2(x_i)/\pi_1(x_i),$$

where π_1 is the value of the post probability associated to the winning class and π_2 is the value of the second highest post probability. In this way, it is possible to judge about the reliability of the classification on the basis of a single value.

Reject Option and Average Cost Minimization

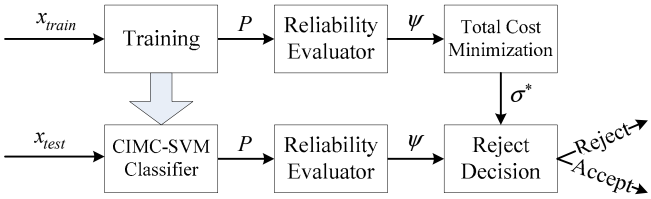

Figure 2. The architecture of the proposed method: the reject option operates on the basis of the reliability evaluator. The optimal value of the reject threshold is established through a training phase.

In classification problems regarding real applications, it should be significant to reject the instances with lowest classification reliability which would otherwise be misclassified [17-19], for the sake of reducing the average cost. Usually, reject means that we have to verify the rejected classification by manpower. So, reject cost R_{reject} is defined as the compensation of the rejection. The reject rule we propose for a CIMCR-SVM compares the classification reliability ψ with a optimal threshold σ^*. The classification is considered acceptable if the reliability value is greater than σ^*, otherwise the input instance is rejected. In fact, the optimal threshold σ^* is computed by minimizing the average cost which can be denoted by:

$$R_{average}(\sigma) = \frac{1}{l}\left(\int_0^{\sigma} R_{reject}\, d\psi(x_i) + \int_{\sigma}^{1} R_{mis}\, d\psi(x_i)\right),$$
$$R_{reject} = \text{constant}, \tag{9}$$
$$R_{mis}^{OAA} = \frac{1}{k}\sum_{s=1}^{k}C_{mis}(t,s),\; R_{mis}^{OAO} = C_{mis}(j,i),$$

where l is the number of training data, R_{reject} is the reject cost function, R_{mis} is the misclassification cost function which denotes as R_{mis}^{OAA} in CI-OAA method and R_{mis}^{OAO} in CI-OAO method. (In practice, the misclassification cost is always bigger than the reject cost.)

The optimal value of the reject threshold σ is the one for which the average cost $R_{average}$ gets its minimum value:

$$\sigma^* = \arg\min_{\sigma\in[0,1]}\left(R_{average}(\sigma)\right). \tag{10}$$

The optimal threshold σ^* can be eventually determined by means of an exhaustive search among the tabulated values of

$R_{average}(\sigma)$.

IV. NUMERICAL EXPERIMENT

A. Evaluation Criteria and Validation

In the fault diagnosis applications, error rate is not an appropriate evaluation criterion when errors have costs. In this paper we use *Average Cost, Accuracy* and *Reliability* as performance evaluation criteria. *Average Cost* is evaluated by (9), while the *Accuracy* and *Reliability* is defined as follow:

$$Accuracy = \frac{T_{accept}}{T_{accept} + T_{reject} + F_{accept} + F_{reject}} \tag{11}$$

$$Reliability = \frac{T_{accept} + F_{reject}}{T_{accept} + T_{reject} + F_{accept} + F_{reject}} \tag{12}$$

where T_{accept} is the number of the correctly classified samples who were accepted by the classifier, T_{reject} is the number of the correctly classified samples who were rejected by the classifier, while F_{accept} is the number of the incorrectly classified samples who were accepted by the classifier, F_{reject} is the number of the incorrectly classified samples who were rejected by the classifier.

In order to reduce the influence of the randomness introduced by the training set extraction process, all the experimental results had been verified by 10 times leave one out cross-validation.

B. Experimental Settings

The experiment data which is get from the vibration test of the steam turbine rotor. There are 80 instances in this data set; each instance has 8 attributes and 1 class label. The characters of the attributes and class labels are described in TABLE I.

In real applications, it is often more expensive to misclassify an actual fault instances into normal, as is shown in TABLE II. That is, the value of $R_{mis}(t,s)$, $s,t \subset \{1,2,3,4\}$, $s \neq t$ is usually larger than that of $R_{mis}(t,s)$, s or $t = 5$, $s \neq t$. When the predict class label of a instance is equal to the actual, the corresponding cost $R_{mis}(t,s) = 0$, $s = t$, that means, the instance is diagnosed correctly. In the applications of fault diagnosis, the losses caused by reject are lower than by misdiagnosis, the reject cost is set as $R_{reject} = 0.8$ with expert experience.

The proposed methods (described in Section 2) have been tested on the diagnosis data set. All the values of attributes in each instance should be normalized between 0 and 1. We first conduct experiments to compare the performance of CI-OAA (cost integrated multi-class classifier use one against all method) and CI-OAO (cost integrated multi-class classifier use one against one method) classifiers with existing standard OAA (multi-class classifier use one against all method) and standard OAO (multi-class classifier use one against one method) classifiers.

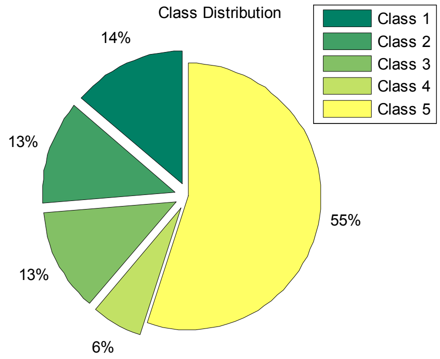

Figure 3. The pie chart displays the class distribution of the diagnosis data

TABLE I. DESCRIPTIONS OF THE CHARACTERS

Characters	Descriptions
f1	(0~0.4)f
f2	(0.4~0.6)f
f3	(0.6~1)f
f4	(1)f
f5	(2)f
f6	(3)f
f7	(4)f
f8	(4~)f
Class(1)	Rotor imbalance fault
Class(2)	Centering fault
Class(3)	Oil whipping fault
Class(4)	Static and dynamic gouging abrasion fault
Class(5)	Normal

TABLE II. MATRIX OF MISDIAGNOSIS COSTS

Actual Class	Predict Class				
	1	*2*	*3*	*4*	*5*
1	0	2	2	2	4
2	3	0	3	3	5
3	6	6	0	6	9
4	7	7	7	0	10
5	1	1	1	1	0

C. Performances and Analyses

In this section we present the results of diagnosis tests aimed at comparing the error-reject trade-off achievable by standard MC-SVM (OAA and OAO) classifiers and by our CIMCR-SVM (CI-OAA and CI-OAO) classifiers. The diagnosis testing procedures is shown in Fig. 1.

1) Average Cost

Average cost is an important indicator in the diagnosis applications, as average cost is directly related to the economic interest. In order to achieve the minimum average cost, a traversal process is executed. In Fig. 4, with the reject threshold increases, the average cost decrease from the maximum to the minimum, while with the reject threshold increases further, the average cost begin to increase contrarily, finally, when the reject threshold increases to 1, all the instances are rejected and the average cost goes to 0.8 ($R_{reject} = 0.8$). That means the employment of reject option make the error-reject trade-off achievable. Comparing the curves in Fig. 4, we can easily find that both OAO and CI-OAO classifiers (with lower average cost and reject threshold) have better performance than OAA and CI-OAA classifiers.

2) Accuracy and Reliability

When the reject threshold is set as 0, that means no samples was rejected. In this condition, the differences between the outputs of CIMCR-SVM classifiers and standard MC-SVM classifiers were only affected by the misdiagnosis cost (shown in TABLE II). Fig. 5 shows the 0-reject classification accuracy of the 5 classes. Standard OAA and OAO classifiers perform worse than CI-OAA and CI-OAO classifiers which are cost sensitive. The sensitivity can be reflected by how much the difference of the numbers of misdiagnosed instances between the standard classifiers and the cost sensitive classifiers is. Comparing with TABLE II, in class 1~4 (fault classes), the higher the misdiagnosis cost, the better the performance. But in class 5, all the instances are predicted correctly with any classifier.

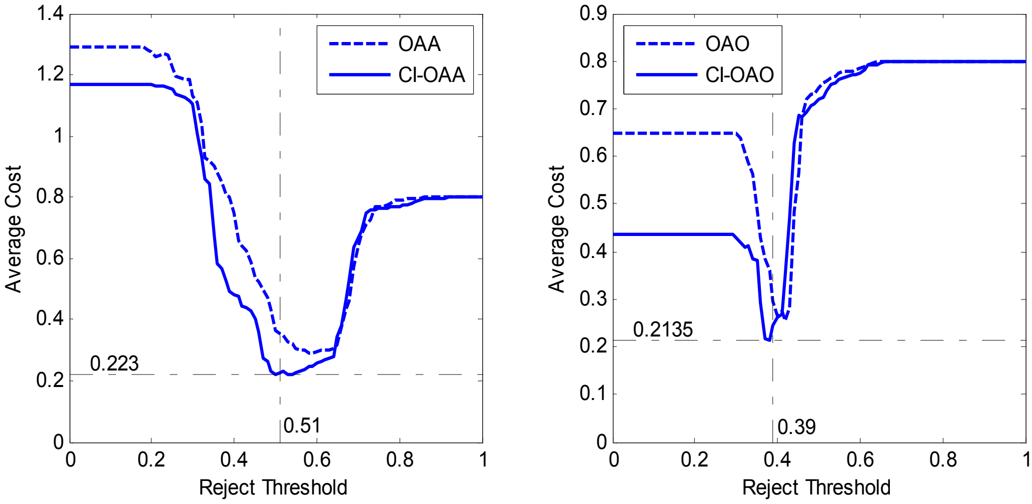

Figure 4. Average cost under different reject threshold

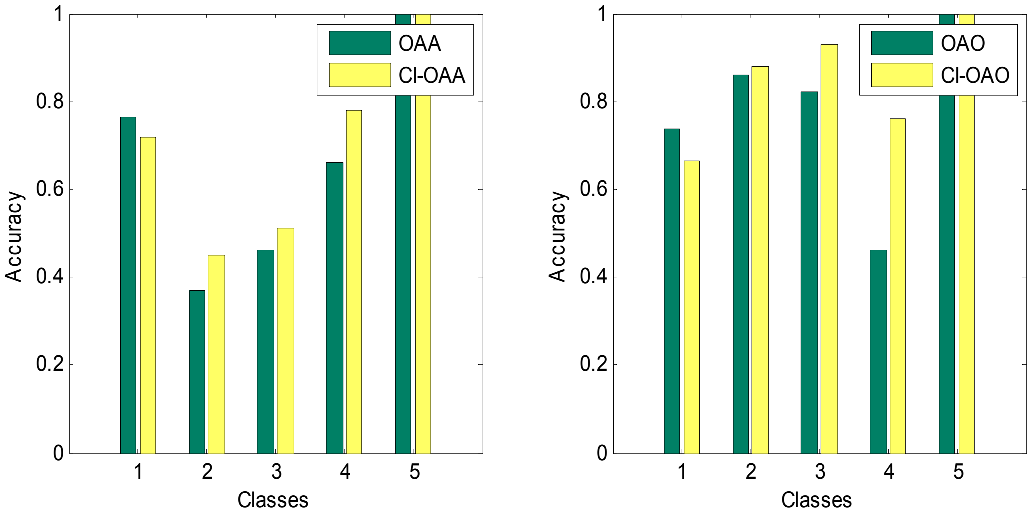

Figure 5. The bar charts display the classification accuracy of the 5 classes when $\sigma^* = 0$.

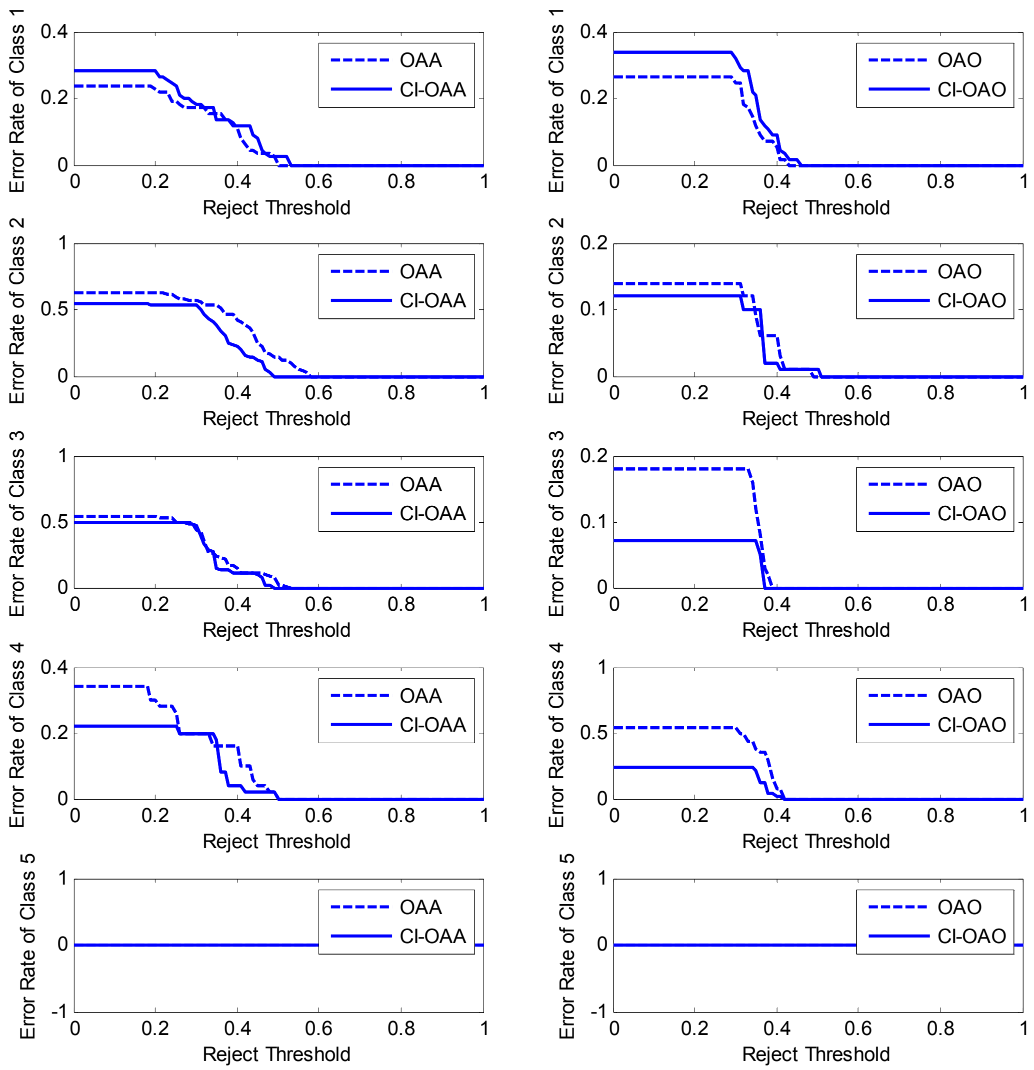

Figure 6. Error rates of 5 classes under different reject threshold

TABLE III. DIAGNOSIS RELIABILITY ON TESTING DATA.

Classes	OAA method				OAO method			
	OAA	**CI-OAA**	**OAA**	**CI-OAA**	**OAO**	**CI-OAO**	**OAO**	**CI-OAO**
	σ=0 (0-reject)		*σ*=0.51*		*σ=0 (0-reject)*		*σ*=0.39*	
1	76.36%	71.82%	100.00%	97.27%	73.64%	66.36%	92.73%	89.09%
2	37.00%	**45.00%**	86.00%	**100.00%**	86.00%	**88.00%**	94.00%	**98.00%**
3	46.00%	**51.00%**	98.00%	**100.00%**	82.00%	**93.00%**	98.00%	**100.00%**
4	66.00%	**78.00%**	100.00%	**100.00%**	46.00%	**76.00%**	70.00%	**96.00%**
5	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

This is no accident, because all the instances of class 5 (normal class) is located in a region of the feature space far from those occupied by the instances of the class 1~4 (fault classes). (This phenomenon also appears in Fig. 6 and TABLE III). This is normally the case in the diagnosis applications. Another interesting conclusion is that the performance of OAO classifiers is better than of OAA classifiers, besides cost sensitive classifiers.

As multi-class classification is more complex process than binary classification, it is meaningful to pay more attention to inner-class relationships. From Fig. 6, we can draw the following conclusions. First, error rates of classes 1~4 (fault classes) decrease when the reject threshold increases. For the purpose of compromising error rates with average costs, we choose the reject threshold with the minimum average cost as the optimal reject threshold. In the fault diagnosis applications, the prediction will be rejected if the classification reliability not reaches up to the optimal reject threshold σ^*. Second, by comparing Fig. 4 and Fig. 6, it is shown that average costs, the minimum average cost and error rates of class 2~4 (fault classes) decrease significantly attribute to the integration of misclassification cost. But this improvement do not take place in class 5 (normal class), because standard OAA and OAO classifiers perform pretty well in class 5 (normal class) already.

TABLE III lists the diagnosis reliabilities on testing data. The diagnosis reliabilities are evaluated by (12). Overall, cost-sensitive multi-class classifiers (include CI-OAA classifier and CI-OAO classifier) achieve better performance than standard multi-class classifiers (include OAA classifier and OAO classifier) on class 2~4. Note that, on class 1, cost-sensitive classifiers perform worse than standard ones. The improvement of diagnosis reliabilities on class 2~4 is at the expense of diagnosis reliabilities on class 1 (class 1 has the lower misdiagnosis cost than class 2~4). Comparing the 0-reject results (when $\sigma = 0$) to the optimal reject results (when $\sigma^* = 0.51$ in OAA methods and $\sigma^* = 0.39$ in OAO methods), the introduction of reject option could markedly improve the diagnosis reliability.

V. CONCLUSIONS AND FUTURE WORKS

In the fault diagnosis applications, it is difficult to recognize the certain types of the faults with higher reliability and lower diagnosis cost. In this works, we propose a fault diagnostic procedure based on cost integrated multiclass SVM with reject option (CIMCR-SVM). The procedure involves two phases, i.e., 1) training a multi-class classifier with OAA or OAO method; 2) rejecting the unreliable diagnosed samples with an optimal reject threshold whose value is obtain by minimizing the average cost. Our experiments show that, comparing with standard multi-class SVM, CIMCR-SVM is able to minimize the average costs (i.e. relate to misclassification cost and reject cost) and improve the diagnosis reliability.

In future work, we will test the method in other fault diagnosis applications. Also, we plan to study the influence of class distribution on classification performances.

REFERENCES

[1] M. Kawada, K. Yamada, K. Yamashita and K. Isaka, “Fundamental Study on Vibration Diagnosis for Turbine Generators Using Wavelet Transform,” Power Systems Conference and Exposition, 2004. IEEE PES, vol. 3, 2004, pp. 1215-1220.
[2] W. M. Lin, C. H. Wu, C. H. Lin and C. H. Su, “Grey Clustering Analysis Based Classifier for Steam Turbine-Generator Fault Diagnosis,” Intelligent Systems Applications to Power Systems, 2007. ISAP 2007. International Conference on 5-8 Nov. 2007, pp: 1-6.
[3] Q. Guo and C. Li, “Research on Data Mining Based on Rough Set and Its Application in Fault Diagnosis of Steam Turbine,” Fuzzy Systems and Knowledge Discovery, vol. 3, 2007, pp. 188-192.
[4] Q. Guo and M. Zhang, “A novel approach for fault diagnosis of steam turbine based on neural network and Genetic Algorithm,” IEEE International Joint Conference on Neural Networks, 2008, pp. 25-29.
[5] C. M. Rocco S. and E. Zio, “A Support Vector Machine Integrated System for the Classification of Operation Anomalies in Nuclear Components and System,” Reliability Engineering and System Safety, vol. 92, 2007, pp. 593-600.
[6] A. Widodo and B. S. Yang, “Support vector machine in machine condition monitoring and fault diagnosis,” Mechanical Systems and Signal Processing, vol. 21, 2007, pp. 2560-2570.
[7] C. Cortes and V. Vapnik, “Support-vector Networks,” Machine Learning, vol. 20, issue 3, 1995, pp. 273-297.
[8] C. W. Hsu and C. J. Lin, "A comparison of methods for multiclass support vector machines," IEEE transactions on Neural Networks, vol. 13, issue 2, 2002, pp. 415-425.
[9] C. Elkan, “The foundations of cost-sensitive learning,” Proceedings of the 7th International Joint Conference on Artificial Intelligence, 2001.
[10] C. X. Ling and V. S. SHENG, “A comparative study of cost-sensitive classifiers,” Chinese Journal of Computers, vol. 30, issue, 8, 2007, pp. 1203-1211.
[11] V. S. Sheng and C. X. Ling, “Thresholding for Making Classifiers Cost-sensitive,” Proceedings of the 21st National Conference on Artificial Intelligence, vol. 1, 2006, pp. 476-481.
[12] N. Abe, B. Zadrozny and J. Langford, “An iterative method for multi-class cost-sensitive learning,” Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2004, pp. 3-11.
[13] X. Y. Liu and Z. H. Zhou, “The Influence of Class Imbalance on Cost-Sensitive Learning: An Empirical Study,” Sixth International Conference on Data Mining, ICDM' 06, 2006, pp. 970-974.
[14] G. Fumera and F. Roli, “Cost-sensitive learning in Support Vector Machines,” In VIII Convegno Associazione Italiana per L’ Intelligenza Artificiale, 2002.
[15] E. H. Zheng, C. Zou, J. Sun and L. Chen, “SVM-based Credit Card Fraud Detection with Reject Cost and Class-dependent Error Cost,” Proceedings of the PAKDD' 09 Workshop: Data Mining When Classes are Imbalanced and Errors Have Cost, 2009, pp. 50-58.
[16] P. Foggia, C. Sansone, F. Tortorella and M. Vento, “Multi-classification: Reject Criteria for the Bayesian Combiner,” Pattern Recognition, vol. 32, 1999, pp. 1435-1447.
[17] C. D. Stefano, C. Sansone and M. Vento, “To Reject or Not to Reject: That is the Question-Answer in Case of Neural Classifiers,” IEEE Trans. On SMC, vol. 30, issue 1, 2008, pp. 84-94.
[18] T. C. W. Landgrebe, D. M. J. Tax, P. Paclík and R. P. W. Duin, “The interaction between classification and reject performance for distance-based reject-option classifiers,” Pattern Recognition Letters, vol. 27, issue 8, 2006, pp. 908-917.
[19] P. Bartlett and M. Wegkamp, “Classification with a reject option using a hinge loss,” Journal of Machine Learning Research, vol. 9, 2008, pp. 1823-1840.
[20] Chao Zou was born in Luoyang, China in 1986. He received his B.S. degree from The First Aeronautic Institute of Air Force, China in 2008.
[21] He is currently pursuing his M.S. degree in China Jiliang University. His research interests include cost-sensitive data mining and machine learning.

En-hui Zheng was born in Xinmin, China in 1975. He received his B.S. and M.S. degree from Northeast Dianli University, China in 1999 and 2002, and Ph.D. degree from Zhejiang University, China in 2006. He is working as an associate professor in College of Mechatronics Engineering, China Jiliang University. His research interests include cost-sensitive data mining and modeling and control of complicated system.

Hong-wei Xu was born in Zhejiang, China in 1972. He received his B.S. degree from Harbin Engineering University, China in 1995, M.S. degree from China Jiliang University, China in 2004. He is working as a lecturer in College of Mechatronics Engineering, China Jiliang University. His research interests include learning control and modeling and control of complicated system.

Le Chen was born in Hangzhou, China in 1959. She received her B.S., M.S. and Ph.D. degree from Zhejiang University, China in 1982, 1994, and 2007. She is working as a research professor and director of Science & Technology Department of China Jiliang University. Her research interests include dynamic measurement, online measurement and temperature metrology.

IACSIT
International Association of Computer Science and Information Technology
WWW.IACSIT.ORG