An Optimized Fault Diagnosis Method for Reciprocating Air Compressors Based on SVM

Nishchal K. Verma
Department of Electrical Engineering
Indian Institute of Technology (IIT) Kanpur
Kanpur, India
nishchal@iitk.ac.in

Abhishek Roy
Department of Electrical and Electronics Engineering
National Institute of Technology Karnataka, Surathkal
Mangalore, India
abhishekroyn@gmail.com

Al Salour
Boeing Company
St. Louis, MO, USA
al.salour@boeing.com

Abstract—**Fault diagnosis in reciprocating air compressors is essential for continuous monitoring of their performance and thereby ensuring quality output. Support Vector Machines (SVMs) are machine learning tools based on structural risk minimization principle and have the advantageous characteristic of good generalization. For this reason, four well-known and widely used SVM based methods, one-against-one (OAO), oneagainst-all (OAA), fuzzy decision function (FDF), and DDAG have been used here and an optimized SVM based technique is proposed for classification based fault diagnosis in reciprocating air compressors. The results obtained through implementation of all five techniques are thus compared as per their accuracy rate in percentages and the performance of the proposed method with 98.03 percent accuracy rate was found to be better than all other classification methods. With the compressor datasets being complex natured, proposed method is found to be of vital importance for classification based fault diagnosis pertaining to reciprocating air compressors.**

Keywords-fault diagnosis; fuzzy decision function; reciprocating air compressor; support vector machine

I. INTRODUCTION

Reciprocating air compressor is one of the key equipments in manufacturing processes, large industrial plants, coal mines, pressurized aircrafts, turbojets etc. Delay in detection of faults in it could cause production loss and product degradation on a large scale and may endanger human life as well. This makes quick and correct fault diagnosis essential for the continuous monitoring of its performance and thereby ensuring quality output [1]. For reciprocating air compressor, occurrence of a fault could result in great economic losses, so the available fault samples in actual fault diagnosis are only a few; hence creating a limiting factor for the implementation of various intelligent fault diagnosis techniques. Vapnik [2] found two main factors which may lead to the failure of ANN model are its insufficient training sample and unreasonable structure design. Support vector machine (SVM) based on statistical learning theory was proposed by Vapnik [2] and is used in many applications of machine learning because of its high accuracy and good generalization capabilities. It is more preferable in classification over artificial neural network (ANN) basically because of using the principle of structural risk minimization (SRM) [3]-[6], rather than using traditional empirical risk minimization (ERM) [7] for classification to minimize the error. Recently SVM has been widely applied for fault diagnosis and classification [8]. Here, apart from proposing an optimized SVM based method, we have chosen four more well-known and widely used SVM based methods, OAO [9], OAA [2],[10], FDF [11], DDAG [12] to detect and classify faults in air compressors. The results obtained through implementation of all five techniques are thus compared as per their accuracy rate in percentages.

This paper is organized into 5 Sections. A brief discussion on the support vector machine is presented in Section-2. In section-3 we review several multi-class SVM methods, such as one-against-one, one-against-all, fuzzy decision function, decision directed acyclic graph and the proposed method. Section-4 discusses about the dataset and presents the results of all the classification methods including the results of comparison for fault diagnosis and classification on the basis of percent accuracy rate obtained. Finally, the conclusions are drawn in Section-5.

II. SUPPORT VECTOR MACHINE

SVMs were designed for binary classifications and its algorithm can be better understood with a mathematical explanation and example as discussed in [13]. Let $S=\{(\mathbf{x}_1,y_1),(\mathbf{x}_2,y_2),...,(\mathbf{x}_l,y_l)\}$ be a training dataset where $\mathbf{x}_i$ are m-dimensional attribute vectors representing feature values, $y_i \in$ {+1, -1}, y_i = 1, and y_i = -1 for class 1 and class 2, respectively. According to [2], the SVMs classifier is represented as

$$D(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) + b = 0, \tag{1}$$

where $\phi(\mathbf{x})$ is a mapping function, $\mathbf{w}^T$ is a vector in the feature space, and b is a scalar. To classify the data which is linearly separable in the feature space, the decision function [13] satisfies the condition

$$y_i(\mathbf{w}^T \phi(\mathbf{x}_i) + b) \geq 1 \text{ , for } i = 1,2,...,l \text{ .} \tag{2}$$

Among all the separating hyperplanes, the optimal separating hyperplane with maximal margin between two classes can be formed by using the condition

$$\min_{\mathbf{w},b} J(\mathbf{w},b) = \frac{1}{2}\mathbf{w}^T\mathbf{w}, \tag{3}$$

subject to (2). If the training data are nonlinearly separable, the hard margin constraints are taken care of by introducing slack variables ξ_i in (2) as

$$y_i(\mathbf{w}^T \phi(\mathbf{x}_i) + b) \geq 1 - \xi_i \text{ , for } i = 1,2,...,l \text{ ,} \tag{4}$$

$$\text{and} \quad \xi_i \geq 0 \text{ , for } i = 1,2,...,l \text{ .} \tag{5}$$

In order to obtain the optimal separating hyperplane, we should apply another condition of minimization [13] as

$$\min_{\mathbf{w},b,\xi_i} J(\mathbf{w},b,\xi_i) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + \lambda\frac{1}{2}\sum_{i=1}^{l}\xi_i \tag{6}$$

subject to (4) and (5), where parameter λ determines the tradeoff between the maximum margin and the minimum classification error. The optimization condition in (6) is a convex quadratic program and can be solved using Lagrange multiplier method. By using Lagrange multipliers α_i and β_i (i = 1, 2,…, l), the Lagrangian function [13] can be constructed as as

$$L(\mathbf{w},b,\alpha_i,\xi_i,\beta_i,) = J(\mathbf{w},b,\xi_i) - \sum_{i=1}^{l}\alpha_i\{y_i[\mathbf{w}^T\phi(\mathbf{x}_i)+b]-1+\xi_i\} - \sum_{i=1}^{l}\beta_i\xi_i \text{ .} \tag{7}$$

According to the Kuhn–Tucker theorem, the solution of the optimization problem can be obtained using Lagrangian function [13] and expressed as

$$\mathbf{w} = \sum_{i=1}^{l}\alpha_i y_i \phi(\mathbf{x}_i) \text{ .} \tag{8}$$

The training examples $(\mathbf{x}_i, y_i)$ having nonzero Lagrangian coefficients α_i are known as support vectors. By solving the following convex quadratic programming problem [13], we can find the α_i coefficients. The problem is formulated as

$$\max_{\alpha_i}\left[-\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} y_i y_j (\phi(\mathbf{x}_i)^T.\phi(\mathbf{x}_j))\alpha_i\alpha_j + \sum_{i=1}^{l}\alpha_i\right] \tag{9}$$

$$\text{subject to} \quad \sum_{i=1}^{l}\alpha_i y_i = 0 \text{ , for } i = 1,2,...,l \text{ ,} \tag{10}$$

$$\text{and} \quad 0 \leq \alpha_i \leq \gamma \text{ , for } i = 1,2,...,l \text{ .} \tag{11}$$

On substituting (8) into (1), the classifier can be obtained. For a new input $\mathbf{x}$, $f(\mathbf{x})$ can be estimated by using (12). If $f(\mathbf{x}) > 0$, the sample is assigned to class 1; otherwise class 2 is assigned to it. The function $f(\mathbf{x})$ [13] is represented as

$$f(\mathbf{x}) = \text{sgn}\{\sum_{i=1}^{l}\alpha_i y_i \cdot (\phi(\mathbf{x}_i)^T \cdot \phi(\mathbf{x})) + b\}, \tag{12}$$

$$\text{where} \quad \text{sgn}(\mathbf{x}) = \begin{cases} 1 \text{ ,} & \mathbf{x} > 0 \\ 0 \text{ ,} & \mathbf{x} \leq 0 \end{cases} \text{ .}$$

In (12), the kernel function [14], [15] is usually used to compute the pairwise inner product in the feature space from the original data items. The kernel function can be represented as

$$K(\mathbf{x},\mathbf{x}_i) = \phi(\mathbf{x})^T \cdot \phi(\mathbf{x}_i) \text{ .} \tag{13}$$

With this, $f(\mathbf{x})$ [13] can be rewritten as

$$f(\mathbf{x}) = \text{sgn}\{\sum_{i=1}^{l}\alpha_i y_i \cdot K(\mathbf{x}_i,\mathbf{x}) + b\} \text{ .} \tag{14}$$

III. MULTICLASS CLASSIFICATION ALGORITHM

In the multiclass classification each of the observations are assigned into one of k classes. In this section, we have briefly introduced the one-against-one, one-against-all, fuzzy decision function, decision-direct acyclic graph method and the proposed method. At first, we discuss the one-against-one algorithm.

Let us assume $S = \{(\mathbf{x}_1,y_1),(\mathbf{x}_2,y_2),...,(\mathbf{x}_l,y_l)\}$ is a training set, where $\mathbf{x}_i \in R^m$ and $y_i \in (1,2,.....,k)$ [13]. For the one-against-one method [9] with k - classes problems, $k(k-1)/2$ classifiers are needed to be determined. The optimal hyperplane [13] with SVMs for class i against class j can be defined as

$$D_{ij}(\mathbf{x}) = \mathbf{w}_{ij}^T \phi(\mathbf{x}) + b_{ij} = 0 ,$$
$$i < j , \quad 1 < j \le k , \quad 1 \le i < k$$

where $\mathbf{w}^T{}_{ij}$ is a vector in the feature space, $\phi(\mathbf{x})$ is a mapping function, and b_{ij} is a scalar. Here the orientation of the optimal hyperplane is represented with the equation

$$D_{ij}(\mathbf{x}) = -D_{ji}(\mathbf{x}) . \tag{15}$$

A. One-Against-One Methods

Given the input vector $\mathbf{x}$, one computes [13]

$$D_i(\mathbf{x}) = \sum_{j \ne i, j=1}^{k} \text{sgn } (D_{ij}(\mathbf{x})), \tag{16}$$

and classifies $\mathbf{x}$ into the class

$$\arg\max_{i=1,\dots,k}(D_i(\mathbf{x})) . \tag{17}$$

B. One-Against-All Method

For a k class problem, the one-against-all method constructs k SVM models. The i^{th} SVM for $i = 1, 2, \dots, k$, is trained with all of the training examples in the i^{th} class with positive labels and all other examples with negative labels. The final output of the one-against-all method is the class that corresponds to the SVM with the highest output value [10]. Thus, by solving the optimization problem in (3)-(5) using all the training samples in the dataset, the decision function of the ith SVM is

$$D_i(\mathbf{x}) = \mathbf{w}_i^T \phi(\mathbf{x}) + b_i , \quad \text{for } i = 1, 2, \dots, k . \tag{18}$$

The input vector $\mathbf{x}$ will be assigned to the i^{th} class that corresponds to the largest value of the decision function, that is, to the class

$$\arg\max_{i=1,\dots,k}(D_i(\mathbf{x})) . \tag{19}$$

C. Fuzzy Decision Function Method

In the FDF method [16], for the input vector $\mathbf{x}$, the one-dimensional membership function [13] $m_{ij}(\mathbf{x})$ for $i, j = 1, 2, \dots, k$, in the direction perpendicular to the optimal separating hyperplanes $D_{ij}(\mathbf{x}) = 0$ is defined as

$$m_{ij}(\mathbf{x}) = \begin{cases} 1 , & 1 \le D_{ij}(\mathbf{x}) \\ D_{ij}(\mathbf{x}) , & \text{otherwise} \end{cases} .$$

The membership functions $m_i(\mathbf{x})$ can be computed [16] as

$$m_i(\mathbf{x}) = \min_{j=1,2,\dots,k}(m_{ij}(\mathbf{x})) , \tag{20}$$

And using (20) classifies $\mathbf{x}$ into the class

$$\arg\max_{i=1,\dots k}(m_i(\mathbf{x})) . \tag{21}$$

D. Decision-Directed Acyclic Graph Method

DDAG method was developed based on the one-against-one scheme [12]. In this a tree type structure is formed representing cases for sample $\mathbf{x}$ regarding which class it could belong to and which class it could not. At the beginning of the tree structure classification, one can choose any pair of class except for the leaf node, and if $D_{ij}(\mathbf{x}) > 0$, then one can consider $\mathbf{x}$ to be not belonging to class j . For example, if there are total three classes for classification and $D_{12}(\mathbf{x}) > 0$, then it means $\mathbf{x}$ does not belong to class 2. It thus belongs to either class 1 or class 3 and the next classification pair is class 1 and class 3. Following the tree structure and repeating similar process, one particular class is obtained for the sample $\mathbf{x}$ at the end of the tree structure and hence the unclassifiable region is also resolved [13].

E. Proposed Method

In the implementation of SVM algorithm using radial basis function (RBF) kernel function, choice of kernel parameter σ plays very vital role in delivering quality performance and obtaining high accuracy value of classification. In the proposed method for a classification based fault diagnosis problem, instead of fixing a particular σ value for the computation of SVM classifier in all cases, the σ value is optimized and its best value is chosen for each and every case whenever SVM classifier is computed. Thus, during implementation of SVM for every individual case, that particular σ value is assigned to it which is found to give highest accuracy rate for that particular case and hence the best σ value may be different for different cases of computation of SVM classifier. Using this method once $D_{ij}(\mathbf{x})$ is calculated, the sample $\mathbf{x}$ is assigned to a class using (16) and (17). Thus, with each individual classification being optimized, the overall performance and classification becomes more effective and better in terms of percent accuracy rate. This could be easily inferred from the table shown in the result section.

IV. DATASETS, RESULTS AND DISCUSSION

The air compressor standard dataset taken here is collected at the workshop, Department of Electrical Engineering, Indian Institute of Technology Kanpur under the Boeing project

TABLE I. REPRESENTATION OF EACH OF THE CLASSES IN THE TRAINING AND THE TESTING DATASETS

Data	Class 1	Class 2	Class 3	Class 4	Class 5	Total
Initial	225	350	350	350	350	1625
Train	113	175	175	175	175	813
Test	112	175	175	175	175	812

Health Monitoring for Rotating Machine. The experiments were conducted on a personal computer with 3.0 GHz CPU and 3 GB of RAM. All the SVM based methods taken were trained by half of the dataset chosen fairly from the main dataset ensuring representation of all the classes present in the required percentage. The remaining half of the main dataset was used for testing and analysis purpose. Table I shows the representation of each of the classes in the training and the testing datasets.

After applying pre-processing techniques, and implementing feature extraction and selection algorithm on the raw dataset, the final dataset on which SVM based methods are implemented contains 1625 rows and 93 columns; here rows describe the instances at which data is acquired and the columns describe various features of the dataset. The dataset taken are basically pressure reading of the compressor in lb/in^2. All together there are 92 feature values in the form of 92 columns in the compressor dataset taken here. There is one more column added to this dataset which represents class of the dataset. *Class* here basically means the state of the compressor system during which the readings are taken from it. As the collective data comprises readings of four different fault conditions and one healthy condition, the data is categorized into five different classes which are illustrated in Table II below along with number of instances belonging to each of the classes in the dataset. We applied the SVM based methods using RBF kernel. The kernel parameter σ and the regularization parameter λ were empirically optimized by minimizing the error rate on the validation dataset. For each problem, we estimate the accuracy rate using different values of kernel parameter σ and regularization parameter λ , where

TABLE II. RECIPROCATING AIR COMPRESSOR FAULTS REPRESENTED AS VARIOUS CLASSES WITH CORRESPONDING NUMBER OF INSTANCES IN THE DATASET

Class	Class Name	No. of instances
Class 1	Non-Returning Valve (NRV) fault	225
Class 2	Leakage outlet valve (LOV) fault	350
Class 3	Leakage Inlet valve (LIV) fault	350
Class 4	Healthy condition	350
Class 5	Unclassified Fault	350

$$\sigma = [2^{-4}, 2^{-3}, 2^{-2}, ..., 2^{7}, 2^{8}],$$

and

$$\lambda = [2^{-4}, 2^{-3}, 2^{-2}, ..., 2^{7}, 2^{8}].$$

In this section, we present the experimental results of implementation of SVM based methods on compressor dataset and compare the performance of the proposed method with the OAO, OAA, FDF and DDAG SVM. Table III and Table IV show and compare the accuracy rates of the proposed method with other methods. It is to be noted that only one optimized value of (σ, λ) is to be needed for computation and while empirically trying to locate their best possible values in terms of yielding highest possible classification accuracy if (σ, λ) show the same highest accuracy rate at more than one values tested, then we highlight and prefer that value of σ and λ which is the least among obtained favorable results.

TABLE III. THE ACCURACY RATE (IN %) W.R.T. σ VALUE

σ	OAO	OAA	FDF	DDAG	Proposed
2^{-4}	26.85	26.85	26.85	26.85	**98.03**
2^{-3}	26.85	26.85	26.85	26.85	
2^{-2}	26.85	26.85	26.85	26.85	
2^{-1}	26.85	26.85	26.85	26.85	
2^{0}	56.90	69.33	54.56	56.90	
2^{1}	96.43	**97.04**	**95.07**	96.43	
2^{2}	56.65	13.79	29.80	77.34	
2^{3}	56.65	13.79	32.02	77.46	
2^{4}	**96.67**	13.79	90.02	**96.67**	
2^{5}	94.83	13.79	89.78	94.83	
2^{6}	92.00	93.84	83.25	92.00	
2^{7}	83.25	90.27	77.71	83.25	
2^{8}	77.34	75.99	75.25	77.34	

TABLE IV. THE ACCURACY RATE (IN %) W.R.T. λ VALUE AT BEST σ VALUE

λ	OAO	OAA	FDF	DDAG	Proposed
2^{-4}	13.79	89.29	71.67	13.79	14.29
2^{-3}	41.50	89.29	71.67	51.11	69.21
2^{-2}	79.06	89.29	81.28	79.06	65.27
2^{-1}	95.07	91.26	**96.43**	95.07	97.04
2^{0}	95.69	**97.04**	95.07	95.69	96.55
2^{1}	**96.67**	97.04	95.07	**96.67**	**98.03**
2^{2}	96.67	97.04	95.07	96.67	98.03
2^{3}	96.67	97.04	95.07	96.67	98.03
2^{4}	96.67	97.04	95.07	96.67	98.03
2^{5}	96.67	97.04	95.07	96.67	98.03
2^{6}	96.67	97.04	95.07	96.67	98.03
2^{7}	96.67	97.04	95.07	96.67	98.03
2^{8}	96.67	97.04	95.07	96.67	98.03

Table III shows the varying accuracy rates in percentage for different σ values. That σ value, at which highest accuracy rate is obtained, will be selected for each of the SVM based methods shown here. Clearly the proposed method gives highest accuracy rate than all other well-known methods taken here. Also, at the chosen optimized σ value for each of the methods, optimized λ value is obtained as shown in the Table IV. FDF is a well known classifier and gives decent accuracy rate of 96.43. One-against-one scores better with accuracy rate of 96.67 percent. One-Against-All comes very closer to the highest accuracy rate with the value of 97.04 but it has a disadvantage of taking relatively higher time in classification. Though DAG shows accuracy rate of 96.67 percent but it was observed to have a big problem of very high time complexity; so notwithstanding its good results it is not a preferred method in this context. With the highest accuracy rate of 98.03 percent the proposed method clearly dominates over all other methods discussed and compared here, and thus it is proposed to be the best method for the classification based fault diagnosis in reciprocating air compressor through SVM based methods.

V. CONCLUSION

Here, classification of the reciprocating air compressor dataset is done through various well-known SVM based methods and the results obtained through the proposed method is found to be dominant over the results of all other methods compared here for classification. Though, DDAG method also gives good results but it has a disadvantage of very high time complexity. Thus, with the optimized selection of RBF parameter σ while calculating kernels in the proposed method of SVM, it could be preferred over all other SVM based methods for the classification of reciprocating air compressor dataset.

The work is on progress to make the proposed method more generalized and robust so that it could be successfully experimented on the problems of various domains to provide quality performance in classification with higher accuracy rate.

ACKNOWLEDGMENT

We are thankful and would like to express our sincere gratitude towards Boeing for allowing us to use its reciprocating air compressor dataset, in our experimental analysis.

REFERENCES

[1] L. Jing and F. Sheng, "Research on remote monitoring system for mining reciprocating air compressor," in Proceedings of the 3rd Congress on Engineering Asset Management and Intelligent Maintenance Systems, WCEAM-IMS, 2008.

[2] V. N. Vapnik, "The nature of statistical learning theory," Springer-Verlag, New York, 1995.

[3] Y. Zhang, X. D. Liu, F. D. Xie and K. Q. Li, "Fault classifier of rotating machinery based on weighted support vector data description," J. Expert Syst. with Appl., vol. 36, no. 4, pp. 7928—7932, 2009.

[4] I. Yélamos, G. Escudero, M. Graells and L. Puigjaner, "Performance assessment of a novel fault diagnosis system based on support vector machines," J. Computers & Chem. Engg., Expert Syst. with Appl., vol. 33, no. 1, pp. 244—255, 2009.

[5] S. Ekici, "Classification of power system disturbances using support vector machines," J. Expert Syst. with Appl., vol. 36, no. 6, pp. 9859—9868, 2009.

[6] N. Wu, X. Liangfa and H. Sanguo, "Fault diagnosis method for power transformer based on ant colony -SVM Classifier," in 2nd International Conference on Computer and Automation Engineering (ICCAE), vol. 1, pp. 629—631, 2010

[7] C. Cortes and V. N. Vapnik, "Support vector networks," J. Machine Learning, vol. 20, pp. 273—293 1995.

[8] F. Sheng, L. Jing and Z. Yabin, "Fault diagnosis system for reciprocating air compressor based on support vector machine," in Proceedings of the 2009 International Workshop on Information Security and Applications (IWISA 2009), Qingdao, China, November 2009.

[9] U. H. G. Kreßel, "Pairwise classification and support vector machines", in Advances in Kernel Methods-Support Vector Learning, B. Schölkopf, C. Burges, and A. Smola, Ed. MIT Press, Cambridge, MA, 1998, pp. 255–268.

[10] R. Debnath, N. Takahide and H. Takahashi, "A decision based one-against-one method for multi-class support vector machine," J. Pattern Anal. Appl. vol. 7, pp. 164–175, Springer-Verlag, 2004.

[11] T. Inoue and S. Abe, "Fuzzy support vector machines for pattern classification," in Proceedings of International Joint Conference Neural Network, vol. 2, pp. 1449–1454, 2001.

[12] J. C. Platt, N. Cristianini and J. S. Taylor, "Large margin DAGs for multiclass classification," in Advances in neural information processing systems, S. A. Solla, T. K. Leen and K. R. Müller, Ed. MIT Press, Cambridge, MA, 2000, pp. 547–553.

[13] B. Liu, Z. Hao and E. C. C. Tsang, "Nesting one-against-one algorithm based on SVMs for pattern classification," IEEE Transactions on Neural Networks, vol. 19, no. 12, 2008.

[14] M. Aizeman, E. Braverman, E. and L. Rozonoer, "Theoretical foundations of potential function method," J. Autom. Remote Control, Pattern Recogn. Learning, vol. 25, pp. 821–837, 1964.

[15] S. Saitoh, "Theory of reproducing kernels and its applications," Longman, Harlow, U.K., 1988.

[16] D. Tsujinishi and S. Abe, "Fuzzy least squares support vector machines for multiclass problems," J. Neural Network, vol. 16, no 5-6, pp. 785–792, 2003.