ANNALS OF FACULTY ENGINEERING HUNEDOARA
– INTERNATIONAL JOURNAL OF ENGINEERING
Tome X (Year 2012) – FASCICULE 3 (ISSN 1584 – 2673)

1. Ashkan MOOSAVIAN, 2. Hojat AHMADI, 3. Meghdad KHAZAEE

COMPRESSOR CONDITION MONITORING BASED ON MULTI-CLASS SUPPORT VECTOR MACHINE

1, 2, 3 DEPARTMENT OF MECHANICAL ENGINEERING OF AGRICULTURAL MACHINERY, UNIVERSITY OF TEHRAN, KARAJ, IRAN

ABSTRACT: *SVM performs machine condition monitoring and diagnosis using its unique ability in classification process. Several methods have been proposed for multi-class classification by combining several binary classifiers. This paper studies the effect of these methods on identification accuracy. Multi-class methods for this work contain of 'one-against-one', 'one-against-all' and 'all-together' with Gaussian kernel. In this work vibration signals for different load conditions were acquired vertically from a compressor on non driven end (NDE). The features are extracted from vibration signal by using statistical parameters that reduce the data and calculation time. These features are used as inputs to the SVM classifier for three class recognition. The experimental results showed that each three multi-class methods were successful in classification with high accuracy. But the performance of 'one-against-all' and 'all-together' methods was better than other method.*
KEYWORDS: *Condition Monitoring; Multi-class SVM; One-against-One; One-against-All; All-Together*

INTRODUCTION

Condition monitoring provides significant information on the health and maintenance requirement of rotary machinery and is used in a vast range of industrial applications [1]. The condition monitoring, diagnostic systems are mainly used to any machines based on vibration and technological parameters measurements [2]. Parameters such as vibration, temperature, lubricant quality and acoustic emission can be used to monitor the mechanical status of equipment. In general, fault diagnosis is a wide and active area of research. There are a large volume of articles that deal with this subject [1].

Most of machinery used in the modern world operates by means of rotary components which can develop faults. The monitoring of the operative conditions of a rotary machine provides a great economic improvement by reducing maintenance costs, as well as improving the safety level. As a part of the machine maintenance task, it is necessary to analyze the external information in order to assessment the internal components state which, generally, are inaccessible without disassemble the machine [3]. One of this information is load condition of machine. Because of it will destroy quickly, if a machine works continuously in full load condition. In addition, the other faults arising in machines are often linked with full load condition. So determining amount of load is significant in life of machine [4, 5].

The Support Vector Machines (SVM) have been developed by Vapnik and are gaining popularity due to many appealing features, and promising empirical performance [6, 18]. SVM, based on statistical learning theory, is a proper technique for solving a variety of learning and function assessment problems. The formulation embodies of SVM is Structural Risk Minimization (SRM) principle [15, 16]. SVM have been successfully applied to a number of applications such as machine condition monitoring, face detection, verification, and recognition, object detection and recognition, handwritten character and digit recognition, text detection and categorization, speech and speaker verification, recognition, information and image retrieval, etc [7, 8]. Support vector machines (SVM) were originally designed for binary (2-class) classification [11]. In binary classification, the class labels can take only two values: 1 and -1. Figure 1 shows the classification of binary SVM.

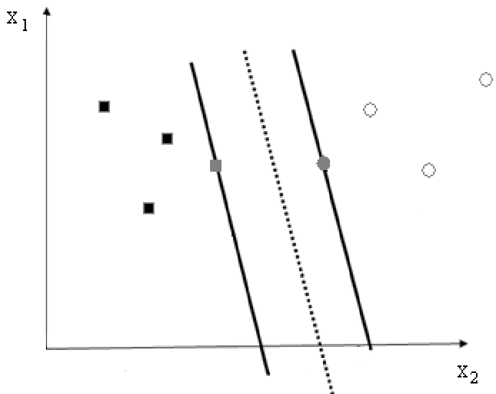

Figure 1. Classification by binary SVM

SVM has been recently applied to solve multi-class problems. Many researches have been done on developing 2-class SVM into multi-class SVM such as [13].

BINARY SVM

Suppose label the training data $\{x_i, y_i\}, i = 1,...,l \quad , y_i \in \{-1,1\}, x_i \in R^d$. *There are some hyperplane that separates the positive (class +1) from the negative (class -1) examples. The vector* x *which lie on the separating hyperplane satisfy* $w.x + b = 0$, *where* w *is normal to the hyperplane. In the separable case, all data satisfy the following constraints:*

$$w.x_i + b \geq +1,\ y_i = +1 \tag{1}$$

$$w.x_i + b \leq -1,\ y_i = -1 \tag{2}$$

These can be combined into the following inequalities:

$$y_i(w.x_i + b) - 1 \geq 0 \quad \forall i \tag{3}$$

$d_+(d_-)$ *is the shortest distance from the separating hyperplane to the closest positive (negative) training data. The margin of a separating hyperplane is defined to be* $d_+ + d_-$. *By constraints Eq.(1) and Eq.(2),* $d_+ = d_- = 1/\|w\|^2$ *and the margin is simply* $2/\|w\|^2$. *Thus we can find the separating hyperplane which gives the maximum margin by minimizing* $\|w\|^2$, *subject to constraints Eq.(3). Using the Lagrange multiplier technique, a positive Lagrange multipliers* $\alpha_i, i = 1,\ldots,l$, *one for each of the inequality constraints Eq.(3) is determined. This gives Lagrangian:*

$$\min \ell_p = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{l} \alpha_i y_i (x_i.w + b) + \sum_{i=1}^{l} \alpha_i,\ \ \alpha_i \geq 0 \tag{4}$$

In order to deal properly with nonlinear SVM, ℓ_p *is transformed into dual problem:*

$$\max \ell_p = \sum_i \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j (x_i.x_j),\ \alpha_i \geq 0,\ \sum_{i=1}^{l} \alpha_i y_i \geq 0 \tag{5}$$

In the case where the data cannot be separated by hyperplane without errors, Vapnik propose that introducing positive slack variables $\xi_i, i = 1,\ldots,l$, *the constraints become:*

$$w.x_i + b \geq +1 - \xi_i \ \text{ for } \ y_i = +1 \tag{6}$$

$$w.x_i + b \leq -1 + \xi_i \ \text{ for } \ y_i = -1 \tag{7}$$

$$\xi_i \geq 0 \tag{8}$$

The goal is to build hyperplane that makes the smallest number of errors. Hence the objection function becomes:

$$\text{minimize } \|w\|^2/2 + C\left(\textstyle\sum_i \xi_i\right) \tag{9}$$

where C *is penalty parameter, a larger* C *corresponding to assigning a higher penalty to errors. The* C *must be chosen by the user. The optimization problem becomes:*

$$\max \ell_D = \sum_i \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j (x_i.x_j),\ 0 \leq \alpha_i \leq C,\ \sum_i \alpha_i y_i \geq 0 \tag{10}$$

Suppose that the data is mapped to a higher dimension space (feature space), using a mapping which is called ϕ

$$\phi : R^d \rightarrow F \tag{11}$$

Then the training algorithm would only depend on the data through dot products in F, *i.e. on functions of the from* $\phi(x_i).\phi(x_j)$. *Kernel function is the significant concept of SVM, The definition of kernel is:*

$$k(x_i.x_j) = \phi(x_i).\phi(x_j) \tag{12}$$

In below, the formulation of three kernels is given:

Linear: $k(x_i.x_j) = x_i.x_j$

Polynomial: $k(x_i.x_j) = (\gamma x_i.x_j + 1)^d,\ \gamma > 0$ 13)

Gaussian RBF: $k(x_i.x_j) = \exp\left(\dfrac{-\|x_i - x_j\|^2}{2\sigma^2}\right)$

which the Gaussian RBF kernel function was applied in the present work [17].
So the optimization problem of nonlinear SVM is:

$$\max \ell_D = \sum_i \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j k(x_i.x_j),\ 0 \leq \alpha_i \leq C,\ \sum_i \alpha_i y_i \geq 0 \tag{14}$$

After solving this optimization problem, those points which $\alpha_i > 0$ are called 'support vectors'. Then they determine w by Eq.(15). And b can be found by KKT 'complementarily' condition Eq.(16), where s_i are support vectors and N_s is the number of support vectors.

$$w = \sum_{j}^{N_s} \alpha_j y_j \phi(s_j) \tag{15}$$

$$\alpha_i (y_j (w.\phi(s_j) + b) - 1) = 0 \tag{16}$$

Finally, the class of x is:

$$sign(w.x + b) = sign\left(\sum_{j}^{N_s} \alpha_j y_j k(s_j, x) + b \right) \tag{17}$$

EXPERIMENTAL SYSTEM

The case study for this work was a gas compressor. The compressor was applied to gas transfer for gas export. The power of compressor was 1360W and it worked on 2979rpm. An accelerometer with vibration analyzer was mounted vertically on the bearing housing of compressor in non driven end (NDE) of shaft to gain vibration signals. The signals of three state of compressor recorded on the memory of data acquisition with sampling rate of 8192 Hz (X-Viber, VMI is manufacturer). The vibration signals of the gas compressor under different load conditions are shown in Figure 2.

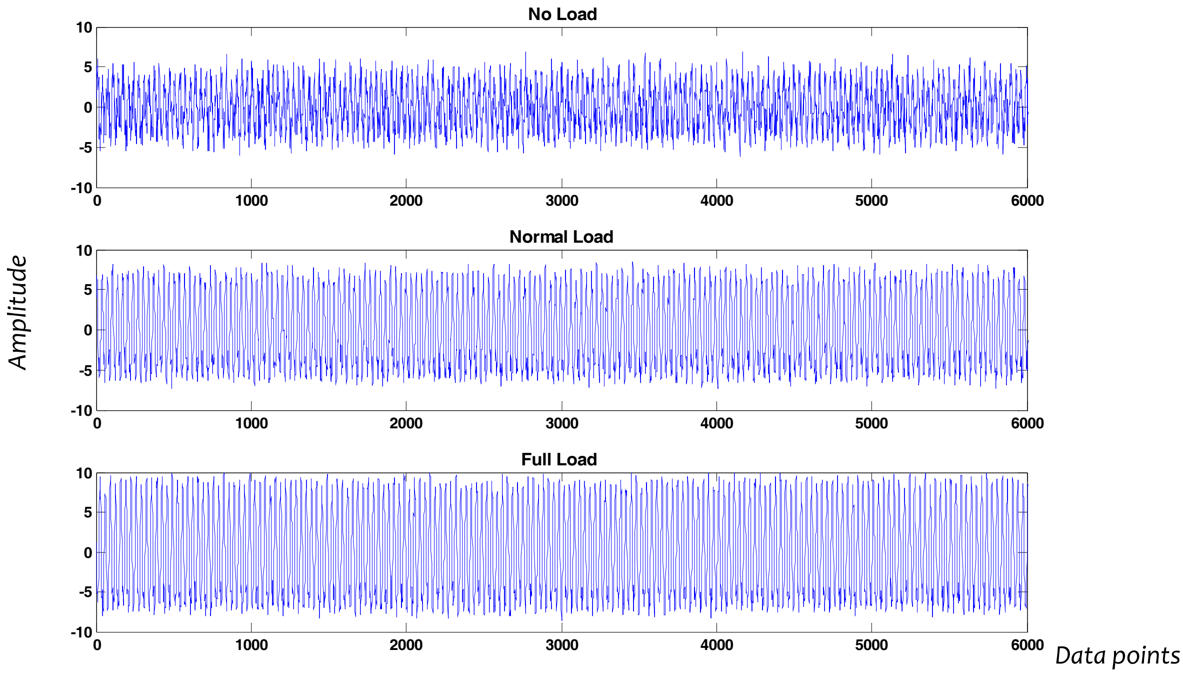

Figure 2. Vibration signals of the gas compressor

Feature Extraction

Vibration signals contain a large set of data for each sample therefore some statistical and frequency domain functions are used to reduce feature vectors. In present work 30 extracted features are used. Some of used parameters are: maximum, minimum, average, root mean square (RMS), standard deviation (Stdv), variance (Var), 5th Momentum, sixth momentum, crest factor, skewness, kurtosis, etc [9, 10].

Multi-class SVM

Support vector machines (SVM) were originally designed for binary (2-class) classification [11, 12, 13]. In binary classification, the class labels can take only two values: 1 and -1. In the real problem, however, we deal more than two classes for examples: in condition monitoring of rotating machineries there are several classes such as mechanical unbalance, misalignment, different load conditions, bearing faults, gear faults, etc. Therefore, in this section the multi-class classification methods will be discussed.

One-against-all (OAA)

The OAA method is earliest implementation for SVM multi-class classification. It builds 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. Thus given l training set $(x_1, y_1), \ldots, (x_l, y_l)$, where $x_i \in R^n, i = 1, \ldots, l$ and $y_i \in \{1, \ldots, k\}$ is the class of x_i, the i th SVM solve the following problem:

$$\text{minimize} \quad \frac{1}{2}\left\|w^i\right\|^2 + C\sum_{i=1}^{1}\xi_j^{\,i}(w^i)^T \tag{18}$$

$$\text{subject to} \quad (w^i)^T\phi(x_j)+b^i \geq 1-\xi_j^{\,i} \quad \text{if} \quad y=i \tag{19}$$

$$(w^i)^T\phi(x_j)+b^i \leq -1+\xi_j^{\,i} \quad \text{if} \quad y\neq i \tag{20}$$

$$\xi_j^{\,i} \geq 0 \ , \quad j=1,\ldots,l \tag{21}$$

where the training data x_i is mapped to a higher-dimensional space by function ϕ and C, is the penalty parameter and ξ, is the slack variable.

Minimizing Eq. (18) means to maximize $2/\left\|w_i\right\|$. When data is not separable, there is a penalty term $C\sum_{i=1}^{l}\xi_{i,i}$, which can reduce the number of training errors [14].

One-against-One (OAO)

This method builds $k(k-1)/2$ classifiers where each one is trained on data from two classes. For training data from the i th and the j th classes, we solve the following binary classification problem:

$$\text{minimize} \quad \frac{1}{2}\left\|w^{ij}\right\|^2 + C\sum_{t}\xi_t^{ij}(w^{ij})^T \tag{22}$$

$$\text{subject to} \quad (w^{ij})^T\phi(x_t)+b^{ij} \geq 1-\xi_t^{ij} \quad \text{if} \quad y_t=i \tag{23}$$

$$(w^{ij})^T\phi(x_t)+b^{ij} \leq -1+\xi_t^{ij} \quad \text{if} \quad y_t=j \tag{24}$$

$$\xi_t^{ij} \geq 0 \ , \ t=1,\ldots,l \tag{25}$$

There are different methods for doing the future testing after all $k(k-1)/2$ classifiers are built. After some tests, the decision is made using the following strategy: if sign $(w^{ij})^T\phi(x_t)+b^{ij}$ says x is in the i th class, then the vote for the i th class is added by one. Otherwise, the j th is increased by one. Then x is predicted in the class using the largest vote. The voting approach described above is also called as Max Win strategy [14].

All-together

In this method, multi-class problems were solved by one single optimization problem [12]. This idea is similar to the OAA method. It builds k two-class rules where the n th function $w_n^T\phi(x)+b$ separates training vectors of the class from the other vectors. Hence there are k decision functions but all are obtained by solving one problem. The formulation is as follows:

$$\text{minimize} \quad \frac{1}{2}\sum_{n=1}^{k}w_n^{\,T}w_n + C\sum_{i=1}^{l}\sum_{n\neq y_i}\xi_i^n \tag{24}$$

subject to

$$w_{y_i}^{\,T}\phi(x_i)+b_{y_i} \geq w_n^{\,T}\phi(x_i)+b_n+2-\xi_i^n \tag{25}$$

$$\xi_i^n \geq 0,\ i=1,\ldots,l\ ,\ \ n\in\{1,\ldots,k\} \tag{26}$$

x is in the class which has the largest value of the decision function $w_n^{\,T}\phi(x)+b_n$.

NUMERICAL EXPERIMENT

Samples for each class divided to two parts, training and testing samples. We give the layout of all data sets in Table 1.

In this work, we studied classification accuracy of each method because of the most important criterion for evaluating the performance of these methods is their accuracy rate. For this comparison, we consider that penalty parameter C of all methods is the same (C=10^3). Also, we use similar condition parameter for QP method (lambda=10^{-7}).

Table 1. Descriptions of data sets in each condition

Class	Train	Test	Total
No Load	45	16	61
Normal Load	45	16	61
Full Load	45	16	61
Total	135	48	183

EXPERIMENTAL RESULTS

For each method, we estimate the train and test accuracy using different kernel parameter (σ):

σ =[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1].

We trained the whole training dataset using the (σ) and predicted the testing dataset. Table 2 shows the classification results on both training and testing dataset. The best accuracy of each method is given.

Table 2. Accuracy rate of three multi-class SVM methods

σ	OAO (Train, Test)%	OAA (Train, Test)%	All-together (Train, Test)%
0.1	(100, 31.30)	(100, 33.33)	(100, 33.20)
0.2	(100, 32)	(100, 35.80)	(100, 34.30)
0.3	(100, 33.33)	(100, 36.42)	(100, 34.89)
0.4	(100, 35.50)	(100, 38)	(100, 37.50)
0.5	(100, 39.2)	(100, 40)	(100, 39.58)
0.6	(100, 45.83)	(100, 47.33)	(100, 46.55)
0.7	(100, 79.17)	(100, 84)	(100, 81.25)
0.8	(100, 89.58)	(100, 91.67)	(100, 90.34)
0.9	(100, 93.75)	(100, 94.67)	(100, 93.75)
1	(100, 95.92)	(100, 98.95)	(100, 98.23)
Best accuracy	(100, 95.92)	(100, 98.95)	(100, 98.23)

We obtained the optimal value of σ and the corresponding classification accuracy for training and testing dataset. We observed that the performance of three methods was similar. But 'all-together' method was really a little slower on the training and testing time. The best accuracy rate of each method was approximately same. In three methods, the accuracy rate was improved continuously by raising the Gaussian kernel parameter (σ). Also the small Gaussian kernel parameter returned the worst accuracy in three methods. Finally, by attention to Table 2, we result that the performance of OAA and 'all-together' methods is quite better than OAO method.

CONCLUSIONS

A comparison between three multi-class methods is presented for diagnosis of different load condition in gas compressor. 29 features consist of statistical and frequency domain parameters have been extracted from vibration signal. These features are performed in support vector machine as input. The efficiency of three multi-class methods, namely, 'one-against-one', 'one-against-all' and 'all-together' has been studied by changing Gaussian kernel parameter (σ). It's be clearly seen that by adding the Gaussian kernel parameter (σ), the performance of each method is better. Also, the best accuracy of each method is 100% on training data and 97.92% on testing dataset.

The results show that the performance of OAA and 'all-together' method is quite better, but 'all-together' method is slower than the other method in classification of training and testing dataset. Also, the result demonstrate that the optimal value of σ is significant in accuracy of classification.

ACKNOWLEDGEMENT

Acknowledgment is made to the especially thanks for University of Tehran for their concentration for this research.

REFERENCES

[1.] B. Huang, Detection of abrupt changes of total lest square models and application in fault detection, IEEE Transactions on Control Systems Technology, 9 (2): 357-367, 2001.

[2.] M. Vasylius et al, The rotating system vibration and diagnostics, Scientific journal Mechanika, 4 (72): 54-58, 2008.

[3.] P.M. Patel and J. M. Prajapati, A review on artificial intelligent system for bearing, International Journal of Engineering Science and Technology, 3 (2): 1520-1525, 2011.

[4.] H. Ahmadi and A. Moosavian, Fault Diagnosis of Journal-Bearing of Generator Using Power Spectral Density and Fault Probability Distribution Function, Innovative Computing Technology, Communications in Computer and Information Science, 241 (1): 30-36, 2011.

[5.] A. Moosavian, H. Ahmadi and A. Tabatabaeefar, Condition monitoring of engine journalbearing using power spectral density and support vector machine, Elixir Mech. Engg, 43: 6631-6635, 2011.

[6.] F. Tian et al, Classification of electronic nose data on wound infection detection using support vector machine combined GA, Journal of Computational Information Systems, 8: 3349-3357, 2012.

[7.] V. Vapnik, An overview of statistical learning theory, IEEE transactions on Neural Networks, 10: 988-1000, 1999.

[8.] C. J. C. Burges, A tutorial on support vector machines for pattern recognition, Data Mining and Knowledge Discovery, 2: 955-974, 1998.

[9.] B. Bagheri, H. Ahmadi and R. Labbafi, Implementing discrete wavelet transform and artificial neural networks for acoustic condition monitoring of gearbox, Elixir Mech. Engg, 35: 2909-2911, 2011.

[10.] B. Samanta, K. R. Al-Balushi and S. A. Al-Araimi, Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection, Engineering Applications of Artificial Intelligence, 16: 657-665, 2003.

[11.] C. Cortes and V. Vapnik, Support-vector network, Machine Learning, 20: 273-297, 1995.

[12.] V. Vapnik, Statistical learning theory, Wiley, New York, 1998.

[13.] J. N. Wang, S. M. Guu and S. T. Chou, On multiclass support vector machines: One-Against-Half approach, International journal of operations research, 6 (1): 24-32, 2009.

[14.] A. Widodo and B. S. Yang, Review support vector machine in machine condition monitoring and fault diagnosis, Mechanical systems and signal processing, 2560-2574, 2007.

[15.] Z. Yin and Y. Li, *Temperature prediction for concrete of bridge structure based on LSSVM and multi-swarm coordination-behaved PSO hybrid algorithm, Journal of Computational Information Systems, 8: 3143-3149, 2012.*
[16.] S. R. Gunn, M. Brown and K. M. Bossley, *Network performance assessment for neurofuzzy data modeling, Intelligent Data Analysis, Lecture Notes in Computer Science, 1208: 313-323, 1997.*
[17.] F. Kuang et al, *A novel approach of KPCA and SVM for intrusion detection, Journal of Computational Information Systems, 8: 3237-3244, 2012.*
[18.] S. R. Gunn, *Support vector machines for classification and regression, Technical report, University of Southampton, Department of Electrical and Computer Science, 1998.*

ANNALS OF FACULTY ENGINEERING HUNEDOARA
– INTERNATIONAL JOURNAL OF ENGINEERING