Premier congrès international
sur l'ingénierie des risques industriels
17-19 décembre 2007, Montréal

First International Conference
on Industrial Risk Engineering
December 17-19, 2007, Montreal

ON THE USE OF TIME SYNCHRONOUS AVERAGING, INDEPENDENT COMPONENT ANALYSIS AND SUPPORT VECTOR MACHINES FOR BEARING FAULT DIAGNOSIS

Komgom N. Christian, Njuki Mureithi, Aouni Lakis, Marc Thomas [1]

Department of Mechanical Engineering, Ecole Polytechnique de Montréal, QC, Canada, H3C 3A7
[1] Department of Mechanical Engineering, Ecole de Technologie Superieure, Montreal, QC, Canada, H3C 1K3
Christian.komgom-nguepjop@polymtl.ca

ABSTRACT

Condition monitoring of rolling elements bearings is investigated in this paper. Recently [11], we have shown that Time Synchronous Averaging combined with Support Vector Machines can lead to efficient bearing fault diagnosis. But the generalization performance of the SVM-boundaries was strongly affected by the transmission path of the signals. This paper is then concerned with the integration of Independent Component Analysis (ICA) in this diagnosis procedure to improve its efficiency in such cases.

First, we validate the use of TSA as a signal processing tool that will automatically highlight bearing defect frequencies if they are present in the envelope spectrum. Next, twenty classical features (rms, peak, crest factor…) are extracted from the envelope of the TSA-signal. To study the influence of Independent Component Analysis on the generalization performance of SVM-boundaries, the twenty dimensional feature vectors are projected in their independent components space. The generalization performance of SVM-boundaries and the influence of signal transmission path as well as the faulty bearing location are then analyzed using these independent components.

1 INTRODUCTION

Condition monitoring of industrial rotating systems has been an active field of research during the past four decades. There is a real demand for reliable procedures that can allow inexperienced users, to detect and characterize any fault condition of an operating system without interrupting its normal running. Condition monitoring is certainly the heart of the ***predictive maintenance*** philosophy. It aims at developing an effective diagnosis procedure that will automatically and accurately detect any fault condition on the running health of a machine. Since any mechanical component, generally progresses through a series of degradation states before failure, if such a running condition can be detected and characterized, then proactive as well as corrective maintenance can be performed before a catastrophic failure occurs. This approach therefore, offers cost savings compared to classical preventive activities which are performed periodically without knowing if the component is really defective.

Among all existing condition monitoring techniques, vibration response analysis is certainly the most used, because it is well understood and it can be performed in many ways [19]. Unfortunately, it has long been recognized that, different applications may require different types of analysis. For example, when rolling elements operate in the presence of strong sources of vibration such as gears, it has been found advantageous to apply high frequency resonance to highlight bearing defect frequencies in the envelope spectrum [12].

Also it is well known that some parameters such as kurtosis may be good indicators of incipient failure of bearings, but become useless at more advanced stage of degradation. So, based on this parameter, bearings may be changed whereas they can still be used without leading to system failure. Recently, the authors [11] have investigated the combination of TSA and SVM for bearing fault detection. We showed that when signals used for training and testing the boundaries are not picked up at the same shaft location, the classification error can reach 40% in most cases whereas the generalization performance is above 90% when training and testing signals have the same transmission path. This paper will now investigate the use of Independent Component Analysis to improve the generalization performance of SVM bounds.

Independent Component Analysis (ICA) as a pre-processing method for feature extraction has been successfully applied in many fields of research ranging from image and speech analysis, electrical brain signals (EEG signal analysis), stock market prediction, telecommunication [6,7,18] to condition monitoring of induction motors [22] and gearboxes [5]. Since ICA should eliminate the effect of random noise, we will determine if it really is the transmission path that affects the generalization performance or if there exists other strong sources of noise in signals that can explain such performance.

Pattern analysis makes it possible to decide whether or not a given bearing is defective or not, just based on its feature vector. Pattern analysis as a decision support tool is well suited for any technician since it does not require strong skills in signal processing techniques. These methods are concerned with the ***minimisation of the risk*** associated to wrong condition state prediction.

First we present an efficient procedure of obtaining TSA signals, from which a 20-dimesional feature vector will be extracted. Next the dimension of these features vectors will be reduced using ICA methods, and the computed independent components will be used in the SVM learning scheme. The generalization performance of the boundaries will now be analyzed and compared to the one previously obtained [11] without ICA.

2   FEATURE EXTRACTION USING TIME SYNCHRONOUS AVERAGING AND INDEPENDENT COMPONENT ANALYSIS

In this section, we present a simplified procedure to obtain TSA signals, and show its results when applied to data collected from the Case Western Reserve bearing data center [23]. Next, we introduce Independent component analysis and its theoretical fundaments. Finally, we explain our feature extraction procedure using these two methods.

2.1 Mc FADDEN & TOOZHY [12] model

Only key points of the analysis are presented here. Interested readers are referred to their paper for details [12]. In this model it is assumed that the envelope spectrum contains peaks only at those frequencies that are harmonics of the characteristic inner race defect frequency f_i surrounded by modulation sidebands at multiples of the shaft rotation frequency f_r given by

$$f = m f_i + n f_r \qquad (1)$$

where m and n are integers. The inner race defect frequency can be related to the shaft rotation frequency by

$$f_i = \frac{z}{2}\left(1 + \frac{D_e}{D_p}\cos\beta\right)f_r = z\left(f_r - f_c\right) \qquad (2)$$

where z is the number of rolling elements, D_e the rolling element diameter, D_p the pitch diameter, β the contact angle and f_c is the cage rotation frequency. So combining (1) and (2) we have

$$f = m z\left(f_r - f_c\right) + n f_r \qquad (3)$$

which shows that some spectrum lines also occur at integer multiples of $f_r - f_c$, the rotation of the shaft relative to the cage. Once the peaks in the envelope spectrum have been identified, we may look for an efficient way to eliminate those that are unwanted in the analysis.

The computation of the synchronous average y(t) of a time signal x(t) using a trigger signal having a frequency f_t is equivalent to the convolution

$$y(t) = c(t) * x(t) \qquad (4)$$

where c(t) is a train of N impulses of amplitude 1/N, spaced at intervals $T_t = 1/f_t$, given by

$$c(t) = \frac{1}{N}\sum_{n=0}^{N-1}\delta\left(t + nT_t\right) \qquad (5)$$

In the frequency domain, this is equivalent to the multiplication of the Fourier transform $X(f)$ of the signal by $C(f)$ the Fourier transform of $c(t)$, i.e.

$$Y(f) = C(f)\cdot X(f) \qquad (6)$$

with $C(f)$ defined as

$$C(f) = \frac{1}{N}\frac{\sin\left(\pi N T_t f\right)}{\sin\left(\pi T_t f\right)} \qquad (7)$$

which is a comb filter. For very large number of averages N, only frequencies at exact multiples of the trigger frequency $f(t)$ are passed. So if we set the trigger frequency to be $f_t = f_r - f_c$, only frequency lines at integer multiples of $f_r - f_c$ will be kept. So, if the synchronous average is calculated using a trigger signal with $f_t = z\left(f_r - f_c\right)$, then the period of the average is $T_t = T_i$, i.e. the period between impacts.

However, computing TSA signals might be too complex for direct implementation in real applications. Its first limitation comes from the fact that a very large number of averages N is required to completely eliminate unwanted frequency lines. So we may be limited to short time duration signal, and the real time trending of a system may be difficult to achieve. Even if this difficulty can be solved with an appropriate trade off between the number of average and the signal duration, the second limitation is certainly the most problematic. It requires getting access to the cage of rolling elements bearings to measure its rotating frequency which is somehow unfeasible in many industrial applications. Facing these difficulties, we look for a simpler procedure to obtain a good approximation of the real time synchronous averaging of raw signals.

2.2 Computation of the time synchronous averaging of vibration response signals: the TSAnalyzer

Time synchronous averaging is a feature extraction technique that have been used successively to gearbox condition monitoring [8]. We have adapted its definition to bearing condition monitoring. TSA-signal is obtained by synchronous averaging of raw signal with a repetitive frequency of the desired signal, or a signal that is synchronous with the desired signal. The residual signal is obtained by removing the primary meshing and shaft components along with their harmonics from the TSA signal.

We validate this approach of computing TSA with signals downloaded form the Case Western bearing data center. These signals come from a test rig made of a 2 hp Reliance Electric motor loaded by a dynamometer. The motor contains faulty bearings that can be mounted at two different locations: drive end (DE) and fan end (FE). Single point faults of different diameters, (7 – 14 – 21 mils; 1 mil=0.001 inches), were created on the bearings separately at the inner raceway, rolling element and outer raceway using electro-discharge machining (EDM). Vibration signals of 10secs were collected at shaft speed of 1797, 1772, 1750 and 1730 rpm, using accelerometers attached on the motor housing at 12 o'clock with magnetic bases.

The TSAnalyzer performs time synchronous averaging and envelope detection of the raw signals by convolving raw signals with a repeating sequence. We can clearly see the importance of TSA as a pre-processing signal tool looking at the power spectral density distributions of figures 1 and 2. Instead of using the shaft rotation frequency as the repeating sequence frequency, we found that it was much more advantageous to use the theoretical bearing defect frequency (namely *BPFO, BPFI, BPFR*). Figures 2 and 3 show how important the choice of repeating frequency is on the envelope spectrum of TSA signal.
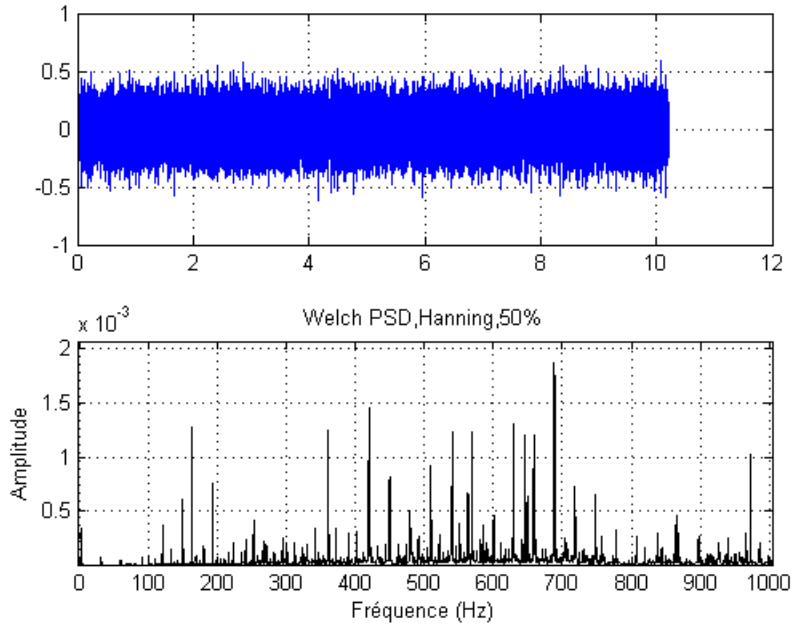
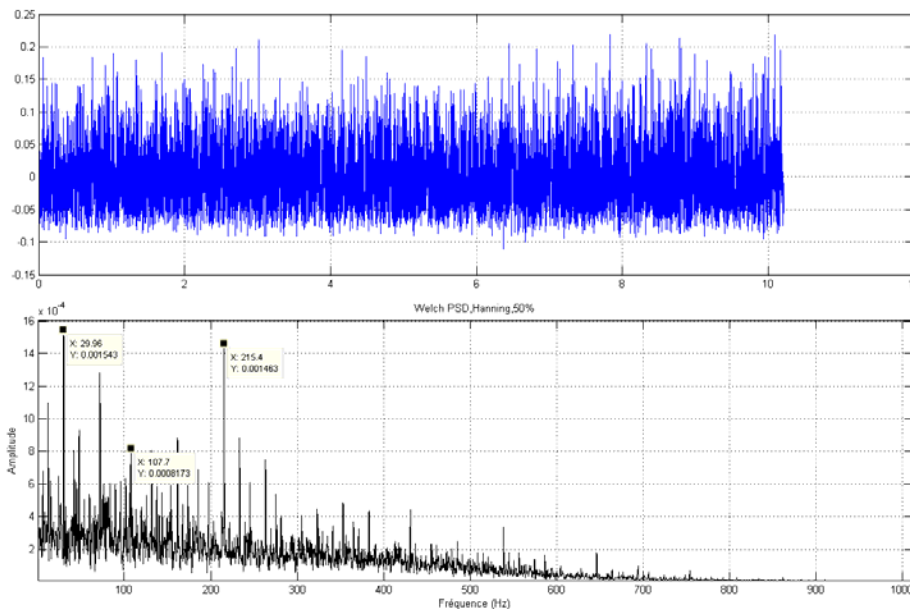*Figure 1 : Power spectral density of raw faulty bearing - 7 mils on the ball - 1797rpm – Not pre-processed by TSA*

Figure 2: Synchronous averaging using shaft rotation frequency as repeating frequency
- 7 mils fault on the ball - 1797rpm

There are many frequency peaks in the spectrum of Fig. 1 compared to the one of Fig. 2 where the shaft rotation (29.95Hz) is the peak frequency, followed by the BPFR second harmonic (215.4Hz). The second harmonic is higher than its fundamental (107.7Hz) because the ball hits the inner and outer race once per revolution of the shaft. Fig. 3 shows clearly that when the BPFR is chosen as the repeating frequency, the envelope spectrum is much freer of undesired frequency peaks. The BPFR and its five following harmonics are well highlighted in the envelope spectrum. Similar results were obtained for faults on the inner (Fig. 4) and outer race

(Fig. 5). The two signals are modulated by the shaft rotating frequency in the case of a fault on the inner race and by the change of the load forces during one revolution when the fault is on the outer race.
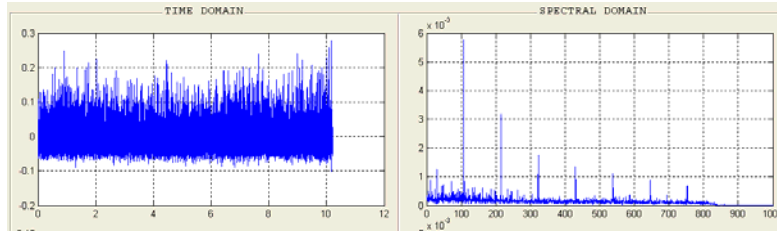

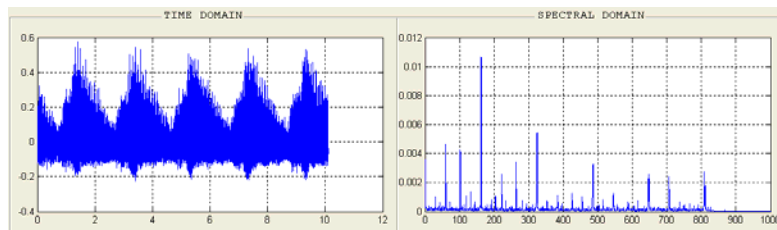*Figure 3: Synchronous averaging using BPFR as repeating frequency - 7 mils fault on the ball - 1797rpm*

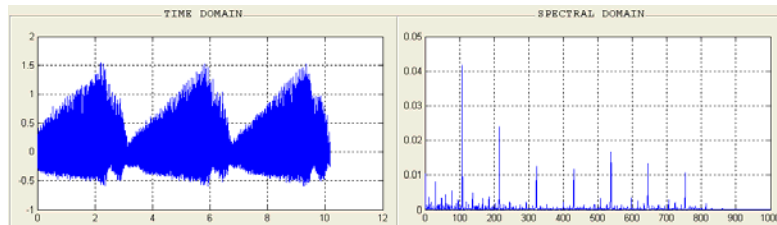*Figure 4: TSA with BPFI, Drive end bearing - 7 mils on the inner race - 1797rpm*

Figure 5 : TSA with BPFO, Drive end bearing - 7 mils on the outer race - 1797rpm

Once the TSAsignals have been computed, we can now compute the residual signal. Fig. 6 shows an example of the residual signal for the case of a 007 mils spall on the inner race. The observed sidebands correspond to the frequencies $f_i - f_r$ and $f_i + f_r$: i.e. 131Hz and 191Hz respectively.
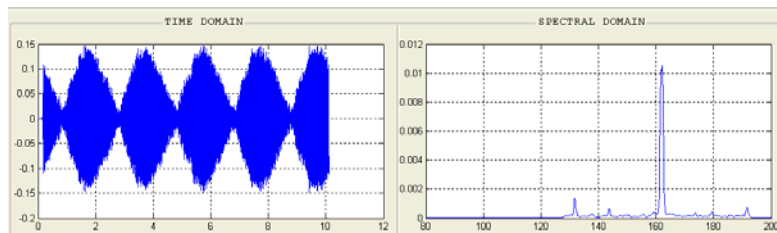

Figure 6: Residual signal, Drive end bearing - 7 mils on the inner race - 1797rpm

2.3   Feature extraction using independent component analysis

2.3.1   Theoretical background

Independent Component Analysis (ICA) is a statistical and computational technique for revealing hidden factors that underlie sets of random variables, measurements or signals. ICA was originally proposed to solve ***Blind Source Separation*** problems. ICA is based on the simple and realistic assumption that if different signals are from different physical processes (e.g. fault on inner and outer race), then those signals are statistically independent. Independent component are found by maximizing a given measure of independence (non-gaussianity) such as kurtosis and negentropy or by minimizing mutual information and maximum likelihood estimation. The fundamental restriction in ICA is that the independent components must be non-gaussian for ICA to be possible.

So given l measured variables, $\mathrm{x}(k) = \begin{bmatrix} x_1(k) & x_2(k) & \ldots & x_l(k) \end{bmatrix}$ at sample k can be expressed as linear combinations of r unknown independent components $\begin{bmatrix} s_1 & s_2 & \ldots & s_r \end{bmatrix}^{\mathrm{T}}$ where $r \leq l$; the relationship between them is given by

$$\mathrm{X} = \mathrm{A\,S} + \mathrm{E} \tag{8}$$

where $\mathrm{X} \in \mathfrak{R}^{l \times n}$ is the data matrix made of n signals, $\mathrm{A} \in \mathfrak{R}^{l \times r}$ is the mixing matrix, $\mathrm{S} \in \mathfrak{R}^{r \times n}$ in the independent component matrix and $E \in \mathfrak{R}^{l \times n}$ is the residual matrix. So the basic problem of ICA is to estimate **S** and **A** only using **X**. One way to achieve this is to calculate a separating matrix **W** so that the components of the reconstructed data matrix S become independent of each other, i.e.

$$\hat{\mathrm{S}} = \mathrm{W} \cdot \mathrm{X} \tag{9}$$

The first step of ICA is to pre-whiten the measured data vector x by a linear transformation, to produce a vector $\tilde{\mathrm{x}}$ whose elements are mutually uncorrelated and all have unit variance. Singular value decomposition (SVD) of the covariance matrix yields

$$\mathrm{C} = \Psi \Sigma \Psi^{\mathrm{T}} \tag{10}$$

where $\Sigma = \mathrm{diag}\begin{pmatrix} \sigma_1 & \sigma_2 & \ldots & \sigma_n \end{pmatrix}$ is a diagonal matrix of singular values and Ψ is the associated singular vector matrix. Then, $\tilde{\mathrm{x}}$ can be expressed as

$$\tilde{\mathrm{x}} = \Sigma^{-1/2}\, \Psi^{\mathrm{T}}\, \mathrm{x} = \mathrm{Q\,A\,s} = \mathrm{B\,s} \tag{11}$$

where **B** is an orthogonal matrix, with $\mathrm{B\,B}^{\mathrm{T}} = \mathrm{I}$. SVD also allows noise reduction by discarding singular values smaller than a given threshold. The problem therefore consists in finding the orthogonal matrix **B** which has fewer parameters than the full rank matrix **A**.

The second step is to employ the fixed-point algorithm to determine W by optimizing some measure of non-gaussianity. Hyvarinen and Oja [6, 7] showed that non-Gaussian is equivalent to independence using the central limit theorem. There are two common measures of non-gaussianity: kurtosis and negentropy. Kurtosis is sensitive to outliers. On the other hand, negentropy is based on the information theoretic quantity of (differential) entropy. Based on approximate form for the negentropy, Hyvarinen [7], introduced a very simple and highly efficient fixed-point algorithm for ICA (available in The FastICA MATLAB Package), calculated over sphered zero-mean vector $\tilde{\mathrm{x}}$. This algorithm calculates one column of the matrix

B and allows the identification of one independent component; the corresponding independent component can then be found as

$$\tilde{x} = B^T Q\, x \qquad (12)$$

The algorithm is repeated to calculate each independent component.

2.3.2 *Feature extraction using ICA*

We have divided each downloaded signals in three equal signals to have more signals available for learning and testing. This way, we obtain 36 data for the inner race fault, 33 for the rolling element fault and 84 for the outer race fault for each bearing location (DE and FE). The feature vector will consist in classical time and frequency parameters evaluated from the TSA signal obtained. As statistical parameters, we have peak value, rms, crest factor, skewness, kurtosis, impulse factor, some statistical moments [10]. As frequency parameters, we take the peak frequency and its amplitude, arithmetic mean, geometric mean, normalized kurtosis, total energy of the spectrum, energy of some frequency interval of the spectrum and FMO [8]. We thus obtain a 20-dimensional feature vector whose independent components will be computed and used by the learning algorithms. The number of independent components was automatically determined by the FastICA algorithm as being 16. The norm of the residual matrix **E** during this process was **7.4**, which is acceptable according to the number of data.

3   SUPPORT VECTOR MACHINES FOR FAULT DETECTION

3.1   Pattern analysis as a principle of risk minimisation

Statistical learning theory was initially developed at the end of sixties by Vladimir Vapnik. Over the four past decades, the learning theory has been successfully applied in many fields; condition monitoring being one of them. Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs) were reported as being the most suitable pattern recognition methods for damage identification in rotating machinery [9, 11]. The main difference between ANNs and SVMs is in the principle of risk minimisation (RM). Statistical learning theory aims in finding hypothesis classes of function defined as $\Phi : x \mapsto f(x)$ which minimize the ***functional risk*** defined by

$$R = \int \frac{1}{2} |y - f(x)|\, dP(x, y) \qquad (13)$$

where x is a set of l elements associated to labels y_i that have an unknown probability distribution function $P(x, y)$. The functional risk is a measure of the error that is made when we are using the hypothesis classes of function Φ to find y. In case of SVMs, structural risk minimisation (SRM) principle is used minimising an upper bound on the expected risk whereas in ANNs, traditional empirical risk minimisation (ERM) is used minimising the error on the training data. The difference in RM leads to better generalisation performance for SVMs than ANNs. Many researchers have reported that SVMs were performing better than ANNs for classification tasks with limited amount of training data.

The Empirical Risk is defined as

$$R_{emp} = \frac{1}{2l} \sum_{i=1}^{l} |y_i - f(x_i)| \qquad (14)$$

Vapnik showed that the empirical risk was an upper bound of the functional risk,

$$R \le R_{emp} + \sqrt{\frac{h\left[\log\left(\frac{2l}{h}\right)+1\right] - \log\left(\frac{\eta}{4}\right)}{l}} \qquad (15)$$

with h being defined as the Vapnik-Chervonenkis (VC) dimension and η the learning rate.

3.2 Support vector classification

Support Vector Machines are a set of efficiently learning systems that use a hypothesis space of linear functions in a high dimensional feature space. The simple SVM algorithm solves a binary problem of classification of data into two distinct classes. The data are divided by a hyperplane which is defined by a number of support vectors. Support vectors are a subset of the training data available for both cases, and are used to define the boundary between the two classes. The use of the support vectors allows complex boundaries to be created, and through the minimization of a quadratic programming problem, the *margin of separation* between each class of data is maximized.

The simplest model of Support Vector Machine is the *maximum margin classifier*. Let N-dimensional input data $x_i, i = 1 \ldots N$ belong to class I or class II, with associated labels $y_i = 1$ for class I and $y_i = -1$ for class II. If the data are linearly separable, we can determine a hyperplane $f(x)$ that will separate the data following the rule $f(x_i) \ge 0$ if x_i belongs to class I and $f(x_i) < 0$ if x_i belongs to class II.

$$f(x) = w \cdot x + b = \sum_{j=1}^{N} w_j x_j + b \qquad (16)$$

with **w** being the N-dimensional normal vector defining the hyperplane and b the learning bias, i.e. the hyperplane distance from the space origin. The optimal hyperplane maximizes the geometrical margin and it can be found by solving the following convex quadratic optimization problem

$$\min \frac{1}{2} \|w\|^2$$
$$\text{s.to } y_i (w \cdot x_i + b) \ge 1 \qquad (17)$$

The corresponding dual problem is then

$$\max W(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,k=1}^{N} \alpha_i \alpha_k y_i y_k x_i x_k$$
$$\text{s.to } \sum_{i=1}^{N} \alpha_i y_i = 0 \text{ and } \alpha_i \ge 0 \qquad (18)$$

One important property of SVMs is that the solution is sparse in α as many training sample lie outside the margin area and their associated lagrangian multipliers α_i are zero. According to the

Karush-Kuhn-Tucker conditions the equality in the constraint of equation (17) will hold for the training pair (x_i, y_i) only if $\alpha_i \neq 0$. In this case, we say that x_i is a *support vector*.
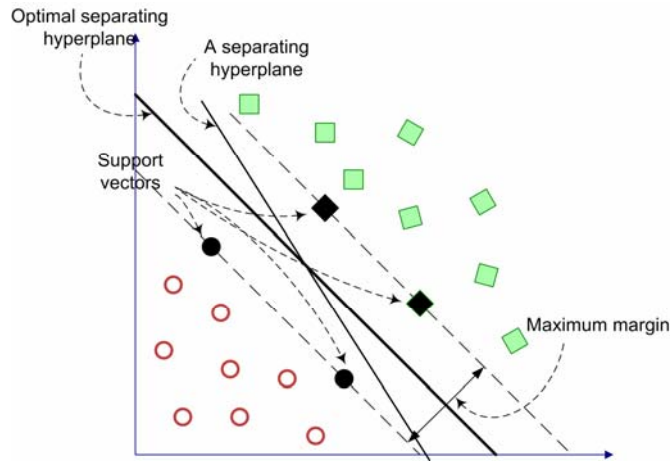

Figure 7 : Illustration of support vectors

If we cannot separate the classes by means of linear decision functions, one can map input vectors onto a higher dimensional feature space F where we expect them to be linearly separable. By choosing a non-linear mapping $\varphi : \Re \to F$ the SVM constructs an optimal separating hyperplane in this higher dimensional space. The hyperplanes in the feature space represent non-linear surfaces and curves in the originating space. Using the non-linear function $\phi(x)$ that maps the N-dimensional input vector x_i into the l-dimensional feature space, the linear decision function is given by

$$f(x_i) = \sum_{i=1}^{N} \alpha_i y_i \phi(x_i) \cdot \phi(x) + b \qquad (19)$$

Fortunately, we can calculate the scalar product $\phi(x_i) \cdot \phi(x)$ without knowing explicitly the mapping function $\phi(x)$ thanks to the kernel trick, i.e.

$$K(x_1, x_2) = \phi(x_1) \cdot \phi(x_2) \qquad (20)$$

The kernel trick exploits the Cover theorem which asserts that a classification problem embedded nonlinearly in a high dimensional space is more likely to be separable than in a low-dimensional space. Making use of the kernel function, the decision function becomes

$$f(x_i) = \sum_{i=1}^{N} \alpha_i y_i K(x_i, x) + b \qquad (21)$$

Many kernel functions (polynomial, radial basis function) are possible, as long as they obey the so-called *Mercer*'s conditions so that equation (20) holds. Several SVMs algorithms were proposed in the literature [2,3,20,21] to handle strong linearity in data.

Multi-class classification using SVMs is still an active research field. The original formulation proposed by [15] uses the *one-against-all* decomposition to a set of binary problems. For *k*-class classification problem, it constructs k binary SVMs, the ith classifier being trained with all the samples from the ith class against all the samples of the rest classes. Another approach for multi-

class classification is the *one-against-one* method. This method consists in training k (k−1)/2 binary SVMs between pair-wise classes. Each classifier casts one vote for its favoured class and the datum to be classified will be assigned to the class which will have the maximum number of votes [6].

In this paper we will opt for the *nu*-soft margin approach using a radial basis function. The optimization of the Lagrangian dual problem is made using the Matlab function quadprog. The multi-classification task will be done using the one against one method, consisting in training 3(3−1) 2 = 3 binary classifiers, i.e.

$$\text{ball vs outer race} = \text{blorBound}$$
$$\text{ball vs inner race} = \text{blirBound}$$
$$\text{inner race vs outer race} = \text{irorBound}$$

Each classifier will be trained and validated with data coming from the drive end fault and tested with data coming from faulty bearing in the fan end fault, in order to test the influence of the signal transmission path on the classification results. For this purpose, we have defined a special notation to handle signals obtained from Drive or Fan end faulty bearing and picked at different locations. For example, a signal obtained from the drive end faulty bearing with a fault on the inner race, and picked up at the fan end side will have the name *DEirfe*. If the fault is on the outer race of the fan end faulty bearing and the signal is picked up at the drive end side, the name will be *FEorfe*. This notation will be used in the presentation of results to specify the nature of the fault, the origin of the signal and the location of the faulty bearing tested.

4   RESULTS

The influence of the faulty bearing location and the signal transmission path on the generalization performance of svm boundaries is measured by testing these boundaries with signals coming from the same and different bearing location (DE and FE) and picked up at different shaft ends.

4.1   Training separately with Drive and Fan End fault bearing signals

Looking at results presented in Tables 1 to 6, we notice that when testing signals come from the same end side (*de* or *fe*), the classification success reaches 95% or higher in general. However, when the testing signals are not collected on the same side than the training signals, the generalization performance is decreased significantly between 70% and 50% in most cases.

Table 1: blirBound, Drive End faulty bearing

Training signals	Testing signals	
	DE...de	DE...fe
DE...de	97.2%	50%
De...fe	50%	98.61%

Table 4: blirBound, Fan End fauly bearing

Training signals	Testing signals	
	FE...de	FE...fe
FE...de	100%	50%
FE...fe	50%	100%

Table 2: blorBound, Drive End fault bearing

Training signals	Testing signals	
	DE...de	DE...fe
DE...de	95.8%	70%
DE...fe	70%	99.2%

Table 5: blorBound, Fan End faulty bearing

Training signals	Testing signals	
	FE...de	FE...fe
FE...de	95.7%	34.7%
FE...fe	64.5%	96.8%

Table 3: irorBound, Drive End fault bearing

Training signals	Testing signals	
	DE...de	DE...fe
DE...de	98.3%	70%
DE...fe	70%	98.3%

Table 6: irorBound, Fan End faulty bearing

Training signals	Testing signals	
	FE...de	FE...fe
FE...de	99%	65.2%
FE...fe	64.5%	99%

The same comments hold for the multi-classification results shown in Tables 7 and 8. Since the training and the testing signals come from the same side, the generalization performance of the classifiers is better than 95% in most cases.

Table 7: Multi-classification success rate using DE bearing -bounds

Boundaries	de-side bounds			fe-side bounds		
Testing signals	DEblde	DEirde	Deorde	DEblfe	DEirfe	Deorfe
Success rate	88.9%	94.4%	97.62%	94.4%	97.2%	98.81%

Table 8: Multi-classification success rate using FE bearing -bounds

Boundaries	de-side bounds			fe-side bounds		
Testing signals	FEblde	FEirde	Feorde	FEblfe	FEirfe	Feorfe
Success rate	100%	97%	93.33%	94%	100%	96.77%

Tables 9 to 14 show that bearing fault location has a more significant impact on the results of the classification. The highest success rate reached is 70% and the lowest one is 30%. These results are not much better than the ones without ICA [11] as expected first. We can explain these results by using the Cover's theorem which asserts that the generalization will always be better in a feature space of higher dimension.

Table 9: blirBound, Drive End faulty bearing

Training signals	Testing signals	
	FE...de	FE...fe
DE...de	50%	50%
DE...fe	50%	50%

Table 10: blorBound, Drive End fault bearing

Training signals	Testing signals	
	FE...de	FE...fe
DE...de	64.5%	65.26%
DE...fe	64.5%	65.26%

Table 11: irorBound, Drive End fault bearing

Training signals	Testing signals	
	FE...de	FE...fe
DE...de	64.5%	65.26%
DE...fe	64.5%	65.26%

Table 13: borBound, Fan End faulty bearing

Training signals	Testing signals	
	DE...de	DE...fe
FE...de	30%	30%
FE...fe	70%	70%

Table 12: blirBound, Fan End fauly bearing

Training signals	Testing signals	
	DE...de	DE...fe
FE...de	50%	50%
Fe...fe	50%	50%

Table 14: irorBound, Fan End faulty bearing

Training signals	Testing signals	
	DE...de	DE...fe
FE...de	70%	70%
FE...fe	70%	70%

4.2 Results of classification when combining Drive and Fan End fault bearing signals when training

In this section, we present the results obtained when signals coming from the two faulty bearings are both used for training. For this purpose, we adopt the following: *debl*, *deir*, *deor*, refer to signals picked up on the drive end side, and *febl*, *feir* and *feor*, for signals picked up on the fan end side.

Tables 15 and 16 give the results obtained when testing and training signals are not picked up on the same side. The generalization performances of the boundaries are the same.

Table 15: Classification success rate when testing de-side boundaries with fe-side signals

de-side boundaries (signals tested)	Generalization
blirbound (*blfe* and *irfe*)	50%
blorbound (*blfe* and *orfe*)	67.9%
irorbound (*irfe* and *orfe*)	67.9%

Table 16: Classification success rate when testing fe-side boundaries with de-side signals

fe-side boundaries (signals tested)	Generalization
blirbound (*blde* and *irde*)	50%
blorbound (*blde* and *orde*)	67.6%
irorbound (*irde* and *orde*)	67.6%

Table 17 gives the success rate in the multi classification scheme. Most stricking results are the **0**% scores obtained by signals with faults on the rolling elements and the inner race when classified by boundaries trained with signals coming from the opposite side.

Table 17: Multi-classification success rate when combining DE and FE bearings during training

Boundaries	*blde*	*irde*	*orde*	*blfe*	*irfe*	*orfe*
de-side	87%	84%	92.3%	**0**%	**0**%	100%
fe-side	**0**%	**0**%	100%	88.4%	88.4%	94.5%

We can explain these results as a particular case of the *one-against-one* method, i.e. each classifier casts a different vote from the other and the data can not be assigned to one class

since each class has the same number of votes. For example *blirBound* votes '**bl**' fault, *blorbound* chooses '**or**' and *irorbound* gives '**ir**', thus each type of faults has 1 vote, and the multi-classifier can not decide which class wins the voting pool. The *one-against-all* method may have performed better in those cases. On the contrary, signals from outer race fault are always well classified even if testing signals come from different sides.

Looking generally at these results, we notice that the SVM boundaries performed better with ICA only in the cases where our first analysis [11] performed well. In others cases, the results are either quite similar or worse. So if ICA is taken as a noise filtration technique as reported in many studies [4-7,18,22], the fact that the generalization performance of SVM boundaries is not better is strictly related to the fact that the data do not contain meaningful information that can be used by the classifiers. This suggests that, the signal transmission path and the faulty bearing location are critical in the development of the diagnosis procedure.

5 CONCLUSION

In this paper, we have studied the integration of independent component analysis in the TSA-SVM based bearing fault diagnosis. Results show that ICA should be handled carefully in this procedure since the generalization performance can be improved only in the case where the first approach has given good results [11]. In fact, the faulty bearing location and the signal transmission path are critical in this analysis, and it may not be realistic to assume that their impact can be attenuated by carrying ICA.

6 ACKNOWLEDGEMENTS

7 REFERENCES

[1] Badri B., Thomas M., Sassi S., Lakis A., Combination of bearing defect simulator and artificial neural network for the diagnosis of damaged bearings, presented at the 20th conference of Condition Monitoring and Diagnostic Engineering Management (COMADEM), June 2007.

[2] Burges C.J.C, A tutorial on support vector machines for pattern recognition, Data mining and Knowledge Discovery, 1998.

[3] Cristianini N., Shawe-Taylor J., An introduction to Support Vector Machines and other kernel-based learning methods, Cambridge University Press, 2000.

[4] Ge Z, Song Z., Process monitoring based on Independent component analysis – Principal component analysis (ICA-PCA) and similarity factors, Industrial & Engineering Chemistry Research, v:46, no:7 : 2054-2063, 2007

[5] He Q., Feng Z., Kong F., Detection of signal transients using independent component analysis and its application in gearbox condition monitoring, Mechanical Systems and Signal Processing 21: 2056-2071, 2007

[6] Hyvarinen A., Karhunen J., Oja E., Independent component analysis, John Wiley & Sons, 2001.

[7]   Hyvarinen A., Oja E., Independent component analysis: algorithms and applications, Neural Networks, 13 (4-5): 411-430, 2000

[8]   Lebold M., McClintic K., Campbell R., Review of vibration analysis methods for gearbox diagnostics and prognostics. Proceedings of the 54th meeting of the Society for Machinery Failure Prevention Technology, 623-634, 2000.

[9]   Lei H., Govindaraju V., Half-Against-Half Multi-class Support Vector Machines, Department of Computer Science and Engineering State University of New York at Buffalo Amherst.

[10]   Krebel, Pairwise classification and support vector machines. In Advances in Kernel Methods: Support Vector Learnings, p 255-268, Cambridge, MIT Press, 1999.

[11]   Komgom N. C., Mureithi N. , Lakis A., Thomas M., Bearing Fault Detection using Pattern recognition methods applied to The Case Western Reserve University Bearing data, accepted for the proceedings of the 25th annual seminar of CMVA October 07, St John, New Brunswick.

[12]   McFadden P.D. and Toozhy M., Application of synchronous averaging to vibration monitoring of rolling element bearing. Mechanical systems and signals processing, 14: 891-906, 2000.

[13]   McFadden P.D., Technique for calculating the time domain averages of the vibration of the individual planet gears and the sun gear in an epicyclic gearbox, Journal of Sound and Vibration, 144 : 163-172, 1991.

[14]   Nandi A.K., Jack L.B., Fault detection using support vector machines and artificial neural networks augmented by genetic algorithms, Mechanical systems and signal processing, v16, no. 2-3, 373-390, 2002.

[15]   Niu X., Zhu L. and Ding H, New statistical moments for the detection of defects in rolling element bearings, International Journal of Advance Manufacturing Technology, 26: 1268-1274, 2005

[16]   Samanta B., Al-Balushi K.R., Al-Araimi S.A., Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection, Engineering applications of artificial intelligence, v 16, n 7-8: 657-665, 2003.

[17]   Scholkopf B., Learnings with kernels: Support vector machines, regularization, optimization, and beyond, Cambridge, MIT Press, 2002.

[18]   Stone J, Independent component analysis, Massachusetts Institute of Technology, 2004

[19]   Tandon N. and Chouldhury A., A Review of vibration and acoustics measurements methods for the detection of defects in rolling element bearing. Tribology Int. Eng., 32: 469-480, 1999.

[20]   Vapnik V., Chervonenkis, A note on one class of perceptrons, Automation and Remote Control, 1964.

[21]   Vapnik V., The nature of statistical learning theory, Springer verlag, 1995.

[22]   Widodo A., Yang B., Han T., Combination of independent component analysis and support vector machines for intelligent fault diagnosis of induction motors, Expert systems with applications, 32: 299-312, 2007

[23]   http://www.eecs.case.edu/laboratory/bearing/