COBEM 99
XV Congresso Brasileiro de Engenharia Mecânica
15th Brazilian Congress of Mechanical Engineering
22 - 26 de Novembro de 1999 / November 22 - 26, 1999 Águas de Lindóia, São Paulo,

USING ACOUSTICAL NOISE FOR FAULT CLASSIFICATION IN GEARBOX

Linilson R. Padovese
Universidade de São Paulo, Escola Politécnica, Dep. de Engenharia Mecânica
Cx. P. 61548 – 05508-900 – São Paulo, SP, Brazil

Abstract. *This paper deals with the development of a computational algorithm for fault classification in gearbox, using acoustical noise. This system was set up to recognise three different patterns independently of the gearbox shaft speed. A Neural Network system is used for this purpose. Two different types of convergence algorithm for the training phase, namely the conjugate gradient and Marquardt, are compared. Concerning the classification task, two different strategies are described and adopted. As pre-processing methods we used a combination of two statistical parameters (rms and kurtosis), and a spectral representation. The results show that is possible to obtain a very high reliability network system to classification purpose.*

***Keywords*:** Neural networks, Acoustics, Fault classification, Gearbox

1. INTRODUCTION

The monitoring of faults occurrence in gearboxes is important for gearbox design and maintenance. In general these faults generate non-stationary multicomponent signals. Several methods of non-destructive monitoring are available for this purpose (vibration, acoustical emission, temperature, etc.). The most popular way of accomplishing this task is vibration analysis. Acoustical noise can also be used for characterisation of defects problems (Braun, 1986).

Non-stationary process can only be correctly analysed by time-frequency (Silva, A. A., 1999, Oehlmann, *et al* 1997) or time-scale (wavelets) methods (Paya, *et al* 1997), but classical spectral representation (Fourier base methods) and statistical methods can also give good representation of the fault signature, depending on the objective of the analysis.

For improving the fault identification and in order to provide an automatic fault classification, neural networks have been widely used in mechanical engineering area, in the last years (Staszweski *et al*, 1997, Schurmann, 1996).

This method can offer an attractive, robust and efficient mean for automation of the classification methodology in conditional monitoring, with applications concerning quality control and predictive maintenance, for example.

An experimental set-up is used to collect acoustical noise data in this study. The two fault patterns used represent local and distributed fault types in spur gears.

The neural network, algorithms of convergence and classification methodology are described. Two kinds of pre-processing methods are used, the statistical and spectral methods.

The results obtained show that simpler pre-processing methods may give excellent results.

2. THE EXPERIMENTAL SYSTEM

The experimental system is composed by an AC motor that drives a gearbox with two reduction spur gear stages, an inverter for shaft speed control, and a Prony brake that generates 60% of the nominal torsion effort specified for the system motor. The faults were produced in the 31 teeth pinion of the first stage, meshing with a 55 teeth gear. These faults were chosen in order to simulate two principal fault sets commonly found in this type of gearbox: partial or total loose of a tooth (local faults) or else contact surface wear in several teeth (distributed faults). In this context three fault pattern types were used: a faultless gear (named normal gear), a tooth-missed gear (named toothless gear) and a gear with pronounced wear in contact surfaces in ten of the teeth (named scratched gear).

The collect data concerns 6 different motor speed values in the interval 400-1400 rpm, with a 200 rpm speed step.

For each fault pattern and shaft speed, an accelerometer (B&K 4393), placed at the vertical direction of the pinion bearing housing, measures vibration signals. At the same time acoustical noise is obtained by a pressure microphone (B&K 4165) placed in front of the pinion location. Both vibration and acoustical signals were low-pass filtered with a cut-off frequency of 2 kHz and sampled with a sampling rate of 5.12 kHz. For each measurement 18 samples of the same signal were obtained. Consequently the database is composed of 324 vibration and acoustical signals.

This database was used to train neural networks as described below.

3. NEURAL NETWORKS

The neural network used in this study is a Multilayer Perceptron (MLP) one. The network characteristics are described in the following items.

3.1 The network topology

An input layer, one or more hidden layers and the output layer form the network topology of a generic MLP. Only the hidden and output are processing layers. The activation function used in the hidden layer is a hyperbolic tangent sigmoid function and the output layer is a linear one. The dimension of the feature vectors or input vectors dictates the dimension of the first layer. The number of elements in the output layer is defined by the dimension of the pattern space or target vectors (dictated by the number of patterns - or faults - to be classified).

Meanwhile the dimensions of the first and last layer are defined by the classification task, there are no rules for determine the hidden layer. Kung (1993) shows that the same results can be achieved by only one layer if an appropriated number of processing neurones is used for unique hidden layer. In the same manner, there is not a unique acceptable rule for establishing the optimum dimension of the hidden layer. Hecht-Nielsen (1990) proposes the maximum number of elements in the hidden layer to be twice the input layer dimension plus one. Maren *et al*. (1990) propose the geometric average between the input and output vectors dimension to be used as the number of elements of the hidden layer, while Baum *et al*. (1989) suggest the dimension VC (Vapnik-Chervonenkis), that depends on an specified generalisation error, for calculating the hidden layer dimension.

Experience shows that in general these rules give only some indication for the hidden layer dimension. In practice, a trial and error method needs to be adopted.

3.2 The learning algorithm

The backpropagation (Kung 1993; Freeman *et al*. 1991) is the learning algorithm used in the MLP. In order to train the MLP to classify any particular set of patterns, it is necessary to have available corresponding pairs of desired inputs and outputs (targets). The input feature presented to the MLP input layer progresses forward through the hidden layers and emerges through the output layer. Then a backpropagation algorithm is used for minimising the sum-squared error between the targets and the outputs calculated by the network. The same procedure is repeated for each element of the input feature set. This is known as the training or learning phase.

There are several methods for accelerating the convergence of the training work, always looking for the minimisation of the sum-squared error. One of then is the popular gradient descent algorithm (Freeman *et al*. 1991), recognised by the learning rate parameter which may also be used with a momentum term to improve the convergence.

Another approach is focused on standard numerical optimisation techniques. The most popular is the conjugate gradient Fletcher-Reeves (CGFR) method (Press *et al*. 1994). This method can be used to minimise objective functions like non-linear least squares functions, that is, the sum squared error function of the network. Hagan *et al*. (1994), proposed a modification for this technique, named Marquardt algorithm (MA) to be incorporated into the backpropagation learning algorithm.

In the present study, we chose to use these last two methods, motivated by their better performance when compared to that of the gradient descent algorithm.

3.3 The training methodology

The goal in the training phase is to obtain a neural network with a good generalisation level. In general the error level controls this characteristic. But a low error level itself doesn't necessarily imply a good generalisation. This error level can be driven to a minimum value on the training phase, but when a new data is presented to the trained MLP, the error may eventually be large. In this case we can say that a good memorisation of the training examples was accomplished, but the network has not learned to generalise new situations.

For this reason, we don't use in this work the error control to establish a good generalisation. Independently of the error level obtained in a training section, the generalisation is controlled by the network performance in recognising a test signal set. For each training task we fixed the number of iteration in the convergence algorithm and then, a set of signal tests was presented to the network. The percentage of hits is our generalisation index.

From the overall signal database, 20% is reserved to this set of tests. The other 80% belong to the training set. Both, training and test databases, were assembled with statistical care, in order to obtain a uniform representation of the parameters used in the experimental phase, namely shaft speed and fault type.

Several factors may contribute to increase the generalisation level. One of them is the network complexity or network dimension, that is, the hidden layer dimension. As mentioned above, a trial and error method was used for this purpose.

Another factor that improves the generalisation level is the fault representation (input features) quality. This can be achieved by convenient signal processing of the input signal. Two kinds of features were used to model the fault signatures in the signal set, as described below.

3.4 The classification strategy

The classification goal is the identification of three patterns of fault independently of shaft speed. For achieving this objective, two approaches were developed, the general strategy and cascade strategy. For the first case, the generic network, only one network was trained to classify the three patterns. This network would have to recognise all the patterns independently of the shaft speed.

For the second strategy, seven networks were used. One of the networks was trained to classify the input features by regarding the shaft speed. The other six were trained to classify the fault patterns, one for each shaft speed. So, after passing through the first network, a signal to be classified would follow into one of the other six, depending on the shaft speed determined by the first network.

Figure 1 illustrates these classification strategies.

The main difference between these two ways is the classification complexity required by the different networks. In the general strategy this complexity is much greater than in the cascade strategy. Consequently the computational effort, topology, dimension of the training set, are greater in general strategy, and the quality of the feature description needs also to be better.

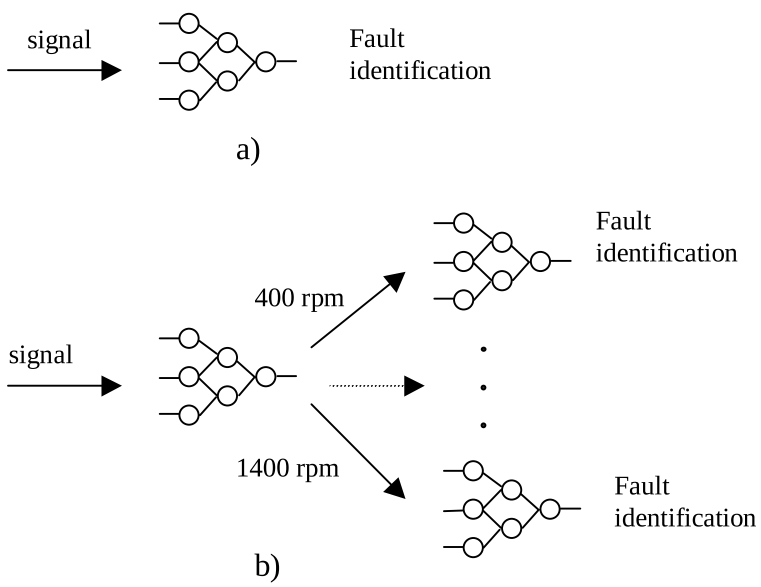

Figure 1 – The two classification strategy: a) the general and b) the cascade classification strategy

4. THE PRE PROCESSING PHASE

The pre-processing phase has a main double objective that is to provide the best feature representation of each fault pattern with the smallest input vector size (to decrease the RNA complexity and computational effort).

Two groups of signal processing methods were used, statistical and spectral analysis.

4.1 The statistical analysis

All random signals s(t), used in this work have the same time duration D, and were centred, that is, they all have zero means values (1^{st}. statistical moment is null).

For each signal, the 2^{nd}., 3^{rd}. and 4^{th}. statistical moments, respectively named *rms*, Skewness and Kurtosis were calculated as well as the probability density function (PDF).

The *rms* of the signal is defined by:

$$rms = \sqrt{\frac{1}{D}\int_0^D s(t)^2\,dt} = \sqrt{\text{Mean Power}} = \sigma \tag{1}$$

This statistical moment can be regarded as a measure of the signal power, as well as its standard deviation σ.

The 3^{rd} and 4^{th} statistical moments are defined by:

$$m_r = \frac{1}{N\sigma^r}\sum_{i=1}^{N} s(t)^r \tag{2}$$

where N is the signal number of points, and the constant r has the value 3 for the 3^{rd} moment and 4 for the 4^{th} moment.

The Skewness, is a 3^{rd}. statistical moment, and it represents a measure of the asymmetry of the signal PDF, in relation to the Gaussian Distribution:

The Kurtosis, K_t, is the 4^{th}. statistical moment, and may be seen as a measure of the flatness or extent of the signal PDF in relation to the Gaussian Distribution:

It was noticed, for rolling bearings that the Kurtosis can be a good fault detection tool for incipient faults and low shaft speed. For the *rms* case it can also be a good fault detection, only for an advanced damage stage (Silva, 1999). Martins & Gerges (1985), and more recently Silva (1999), proposed a combination of these two parameters, K_t and *rms* , with the objective of meshing both statistical qualities, given by:

$$K = K_t \cdot rms \tag{3}$$

Figure 2 shows the *rms*, K_t and K parameters for the three fault patterns (normal, toothless and scratched gear) as function of the shaft speed. For each curve it is indicated the standard deviation of the parameter, calculated over 54 signals collected for each fault pattern.

As it can be seen in Fig. 2, the K parameter allows separating the three fault patterns into three distinct classes. The Kurtosis has a multiplier effect over the *rms*.

Figure 3,presents an example of a PDF of the acoustical signal for the three pattern faults, at a shaft speed of 1000 rpm. Each PDF was normalised keeping the area under each curve equal to unit. This graphical pattern is analogous to PDFs coming from different shaft speeds. In all them, the scratched gear PDF pattern is very different from the others. The normal and toothless gears are similar in shape.

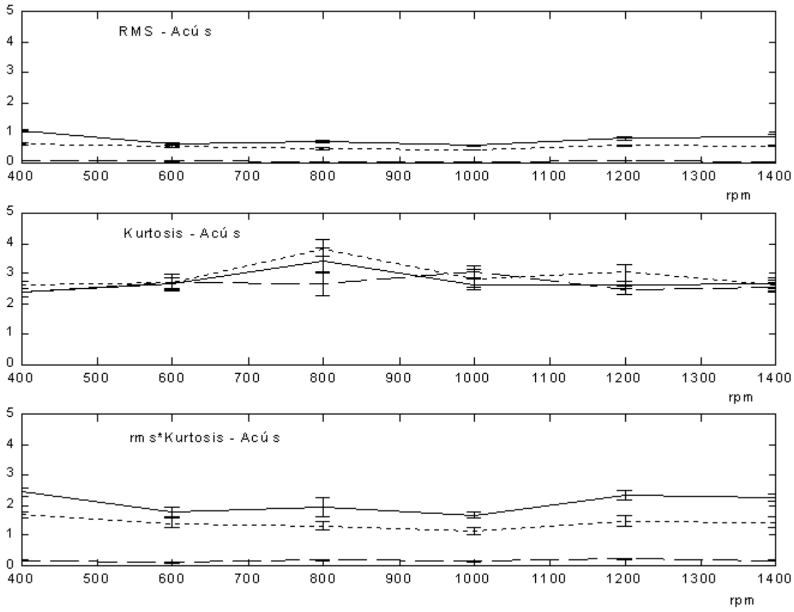

Figure 2 – *rms*, K_t and *K* for acoustical signals coming from three fault patterns: normal (solid line), toothless (dashed line) and scratched gear (dotted line) as function of the shaft speed.

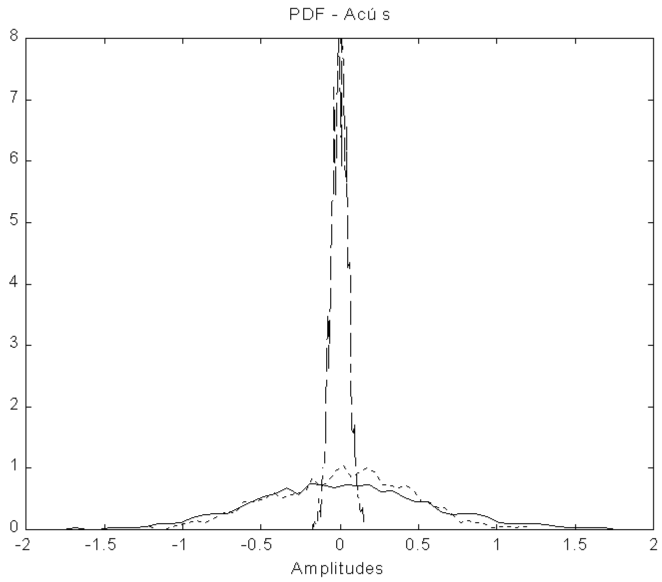

Figure 3 – PDF of the acoustical signals from three faults pattern: normal (solid line), toothless (dot line) and scratched gear (dashed line), for a shaft speed of 1000 rpm.

4.2 The spectral analysis

The Power Spectral Density (PSD) of the acoustic signals was obtained with the Welch method (Proakis, Manolakis, 1996). Each spectral representation is a result of an average of 10 spectra.

Figure 4 shows 3-dimensional graphics for each fault. The PSD is plotted for each one of the 6 shaft speeds. The scale bar, on the right side, is an amplitude scale for the PSD representation.

In all of the three graphics it's possible to distinguish two classes of spectral patterns. One of them is seen as a constant spectral pattern do not changing with the shaft speed. The other spectral pattern depends on the shaft speed in a linear way, increasing with decreasing shaft speed. There isn't a net distinction among the three spectral representations.

As it seen in Fig. 2 (*rms* graphic) the maximum power signal occurs for the normal gear, then, in a decreasing way, for the toothless gear, and finally, the lowest power signal corresponds to the scratched gear. Regarding to Fig. 4, at the maximum of the amplitude scales, it is possible to observe this same pattern. The square root of the integration of each signal over the frequency domain results in the same mean power estimation, or *rms,* as defined by Eq. (1).

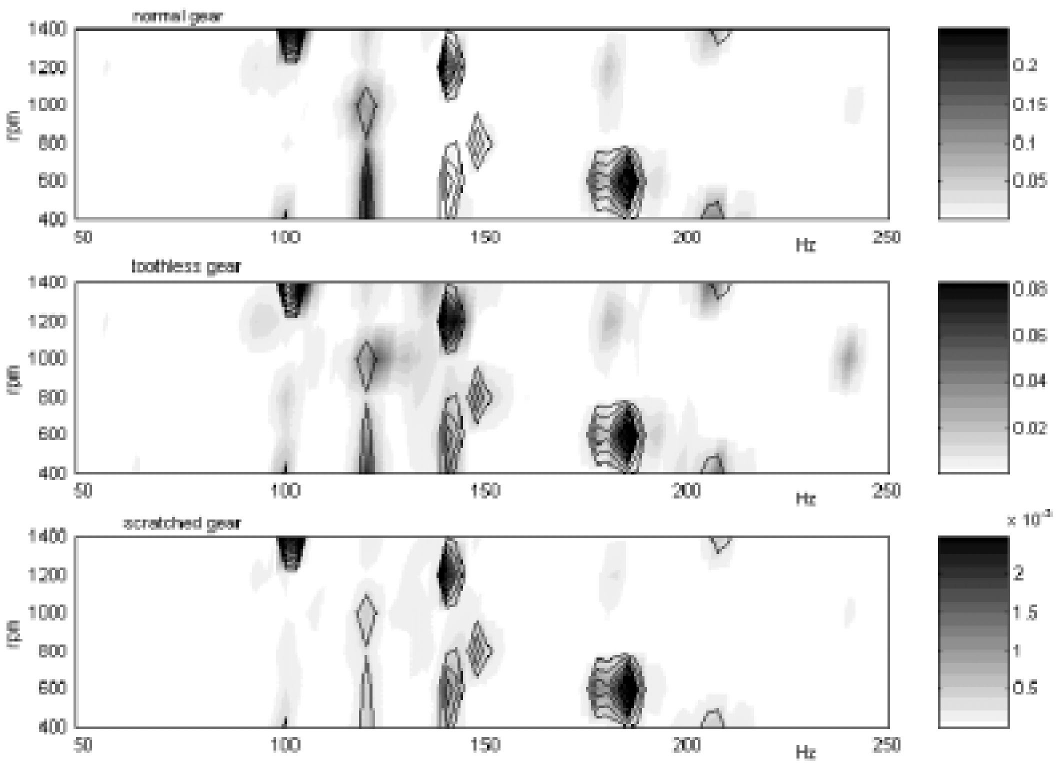

Figure 4 – Graphical 3D representation of the acoustical spectra against the shaft speed variation.

The input spectral vectors have 81 points, with a frequency band of 50 - 250 Hz, and sampling frequency of 512 Hz. The frequency resolution is 2.5 Hz.

5. RESULTS AND DISCUSSION

For neural networks training purposes two input features vectors are used: the *K* parameter and the spectral representation.

The number of interactions in all training sections was set to 500. For all results presented below, the sum-squared error was smaller than 10^{-2}.

Table 1 shows the principal results of the generic network training. The first column presents the input features used. The Skewness was also utilised, but as the results obtained were very poor they are not mentioned in this text.

Table 1. Results for the generic network.

Input feature	Network topology	Best generalisation [worst Q]	Convergence algorithm
K	1 x 10 x 2	90% (48/54) [37%]	CGRF good MA good
Spectra	81 x 20 x 2	100 % (54/54) [21%]	CGRF good MA bad

The topology with which the better results were obtained is presented in the second column. Several dimensions for the hidden layer were used, in a trial and error method. Then the presented topology may not be optimum. The output dimension was obtained by codification of the three fault patterns, that is (0.5, 0.5) for the normal gear, (-0.5, 0.5) for the toothless gear and (-0.5, –0.5) for the scratched gear. Consequently, the three target patterns are distributed in three quadrants.

The best results obtained for each one of the input features are given in the third column. It is also indicated in parenthesis, the number of hits over the dimension of the database test.

It is assumed that if a calculated pattern falls in the correspondent quadrant, it is correctly classified; and its classification quality is evaluated by the distance between the code calculated by the network and the target code, given by:

$$Q = \frac{\sqrt{(\bar{a} - c_1)^2 + (b - c_2)^2}}{\sqrt{c_1^2 + c_2^2}} \times 100 \tag{4}$$

where (a, b) is the calculated code, and (c_1, c_2) is the correspondingly target code. The parameter quality Q represents the distance in percentage between the target and the calculated code. It is also presented in brackets in the third column, the worst value for the parameter Q.

Regarding to Fig. 2 and Fig. 4, it seems that the K parameter is a better feature, but the results presented in the third column show the opposite. The spectra feature gives the best result.

In column four, it is indicated the performance of two convergence algorithm: the CGRF, conjugate gradient algorithm and the MA, Marquardt algorithm. Fast convergence was obtained with both as well as good level of generalisation, for the network of small complexity. For bigger networks, the timing convergence in the case of the MA algorithm is very low, and the results worse.

The gradient descent algorithm was also used for performance comparison. As it is a parametric method (it is necessary to give the learning rate and momentum parameter) its performance is very low when compared with the CGRF and MA methods.

Table 2 shows the principal results for the first network layer of the cascade strategy.

For shaft speed classification the best input feature was the spectra representation. It was not possible to obtain any reasonable result with the K parameter. The reason is that this parameter doesn't provide class distinction, for the different shaft speeds. Figure 2 shows this fact. Again, the MA algorithm gave worse results when compared to the CGRF algorithm, for the spectra input feature case. A codification of dimension three was used in this network.

Table 2. Results for the first network layer of the cascade network system.

First Network Shaft speed classification			
Input Features	Network Topology	Best Generalisation [Q]	Convergence Algorithm
K	1 x *n* x 3	Not occur	-
Spectra	81 x 20 x 3	98% (53/54) [32%]	CGRF good MA bad

Table 3 resumes the results of the training task for the six networks of the cascade strategy (one for each shaft speed). Each of them was trained to recognise the three fault patterns. For each type of input features, all the six networks have the same topology. Again, the best results were obtained with the spectra feature. It was obtained 100% of hits for all the six networks, and again the performance of the MA algorithm decrease with the increasing of the network complexity.

Table 3. Results for the second network layer of the cascade network system.

Second Network Fault classification for each shaft speed			
Input Features	Network topology	Best and worst Generalisation [Q]	Convergence Algorithm
K	1 x 5 x 2	100 %– 94% (18/18) – (17/18) [46%]	CGRF good MA good
Spectra	81 x 10 x 2	100 % (18/18) [37%]	CGRF good MA bad

6. CONCLUSION

The results show that it is possible to obtain a very high reliability classification system to diagnose fault types in a gearbox, using acoustical noise. Environmental noise added to the acoustical noise emitted by the gearbox can degrade this performance, and this effect needs to be studied yet.

The input feature type is determinant for a good network generalisation. The methods used to pre-process the data, *K* parameter and spectra representation, give good results with a better performance for the spectra representation.

The choice of the convergence algorithm is also important. Three algorithms were used in this study. The gradient descent is a parametric algorithm. Two parameters have to be given in this case, the learning and momentum parameters. The performance was the worst for this algorithm. In the case of small network complexity the performance of the two others (the conjugate gradient and the Marquardt algorithms) were comparable. But for networks with a higher complexity the performance of MA algorithm is very poor.

The classification strategy can also influence the network characterisation. The cascade strategy gave better results, in generalisation level. Its smaller classification complexity contributes for this result.

Acknowledgements

The author wishes to thank FAPESP for the support given for this work through Project no. 94/3538-9.

REFERENCES

Baum, E.B.; Haussler, D.; 1989, What size net gives valid generalization?; Neural Computation, Vol 1, pp 1151-160.

Braun, S.; 1986, Mechanical Signature Analysis – Theory an Applications. Academic Press, New York.

Freeman, J. A.; Skapura, D. M.; 1991, Neural Networks. Algorithms, Applications, and Programming Techniques; Addison-Wesley.

Hagan, M. T.; Menhaj, M. B.; 1994, Training Feedforward Networks with the Marquardt Algorithm, IEEE Transactions on Neural Networks, Vol. 5, No. 6, November.

Hecht-Nielsen, R.; 1990, Neurocomputing; Addison-Wesley.

Kung, S.Y.; 1993, Digital Neural Networks; Prentice Hal; New Jersey.

Maren, A. J.; Harston, C.T.; Pap, R.M.; 1990, Handbook of Neural Computing Applications; Academic Press, San Diego.

Martins, L. G., Gerges, S.N.Y.; 1985, Manutenção preventiva: um estudo sobre o método da curtose, COBEM, 8th, São Paulo, p.673-676.

Oehlmann, H., Brie, D., Tomczak, M., Richard, A.; 1997, A Method for Analysing Gearbox Faults using Time-Frequency Representation, Mechanical System and Signal Processing, vol. 11, No. 4.

Paya, B. A., Esat, I. I., Badi, M. N.; 1997, Artificial Neural Network Based Fault Diagnostics of Rotating Machinery Using Wavelet Transforms as a Pre-processor; Mechanical System and Signal Processing, Vol. 11, No. 5.

Press, W. H.; Vetterling, W. T.; Teukolsky, S. A.; Flannery, B. P.; 1994, Numerical Recipes in C; 2nd. Ed., Cambridge University Press.

Proakis, J. G.; Manolakis, D. G.; 1996, Digital Signal Processing, 3rd. Ed., Prentice Hall.

Schurmann, J.; A.; 1996, Unified View of Statistical and Neural Approaches; John Wiley & Sons.

Silva, A. A.; 1999, Detecção e Análise Dinâmica de Falhas em Rolamentos; Tese de Doutorado, EESC-USP.

Staszewski, W.; Worden, K.; Tomlinson, G. R.; 1997, Time-Frequency Analysis in Gearbox Fault Detection Using the Wigner-Ville Distribution and Pattern Recognition; Mechanical Systems and Signal Processing, Vol. 11, No. 5.