IMPLEMENTATION OF THE HYBRID PARADIGM OF WEIGHT'S INITIALIZATION FOR APPLICATION IN ARTIFICIAL NEURAL NETWORKS MULTILAYER

ABST2071

Abstract. *Different factors influence in the speed of supervised training of artificial neural networks (ANN's) multilayers and in the quality of the results after the convergence. Among them they are distinguished the ANN's topology, algorithm of learning, function of activation of the neurons and, mainly, the initial set of sinaptics weights. In this work a simple and robust alternative for initialization of weights for application in artificial neural networks multilayers through the method of the hybrid paradigm is shown. The method of the hybrid paradigm for initialization of weights can be interpreted as intermediate between the paradigms of the easiest way and the shortest way that are alternative paradigms for search of the best initial joint of weights. The developed methodology is applied to a set of four excitations (2, 3, 5 and 7 broken bars) beyond the normal condition of functioning. The deterministics frequencies of broken bars through collected spectra of vibration will be shown experimentally. Through a selective filter it is possible to reduce the number of parameters capable to represent the signals used for training of the ANN's. The behavior will be observed and presented a comparison between the results of training and validation of ANN's (broken bars and normal condition of functioning) through the random initialization method and of the hybrid paradigm method.*

Keys words*: Artificial Neural Networks, Synaptics Weights, Method of the Hybrid Paradigm*

1. Introduction

The process of initialization of the sinaptics weights among others factors such as the architecture of the net, the function of activation of the neurons and the rule of learning have great influence in the quality and efficiency of the supervised training of artificial neural networks (ANN's) multilayers. The training can evolve more quickly if an initial adjusted choice for the values of the weights can be done. On the other hand problem of the net's saturation known as paralysis of the training or the convergence for poor local minimums can occur.

Currently there are several techniques with proposals that define the set of initial weights. The simplest methods are based on a random distribution uniform, Kolen and Pollak (1990). These techniques are still sufficiently employees mainly due the facility of implementation and the flexibility of configuration. However the period of training is a long way, Uliana *et al*. (1997). On the other hand, to abbreviate the training process the no random initialization method tries to initialize the weights next to the best solution. In some applications the necessary additional process becomes them well slower of the one than the random initialization. The adjusted choice for the initialization of the weights can be finding between great and small values, Haykin (2001).

For the process of resultant optimization, there are two alternative paradigms for searching of the best initial set of weights, the paradigm of the easiest way and the paradigm of the shortest way, Castro and Von Zuben (1998).

In this work, a simple and robust alternative for initialization of weights for application in artificial neural networks multilayers through the method of the hybrid paradigm is shown.

The developed methodology is applied to a set of four excitations (2, 3, 5 and 7 broken bars) beyond the normal condition of functioning. The deterministics frequencies of broken bars through collected specters of vibration will be shown experimentally. Through a selective filter it is possible to reduce the number of parameters capable to represent the signals used for training of the ANN's.

The behavior will be observed and presented a comparison between the results of training and validation of ANN's through the random initialization method and of the hybrid paradigm method.

2. Experimental tests

The experimental tests have been done in the Federal University of São João del Rei. The instrumented test desk is shown in Fig. 1. The three phases, induction motor, squirrel-cage rotor, 5 HP, 220V, 60 Hp, 4 poles, 44 bars, 36 slots, 1730 rpm nominal rotation, have been powered with phase balance voltages. To simulate load, it has been used a CC generator [1] that power a resistance bench [2] linked to an electrical motor [3] through a flexible linking [4].

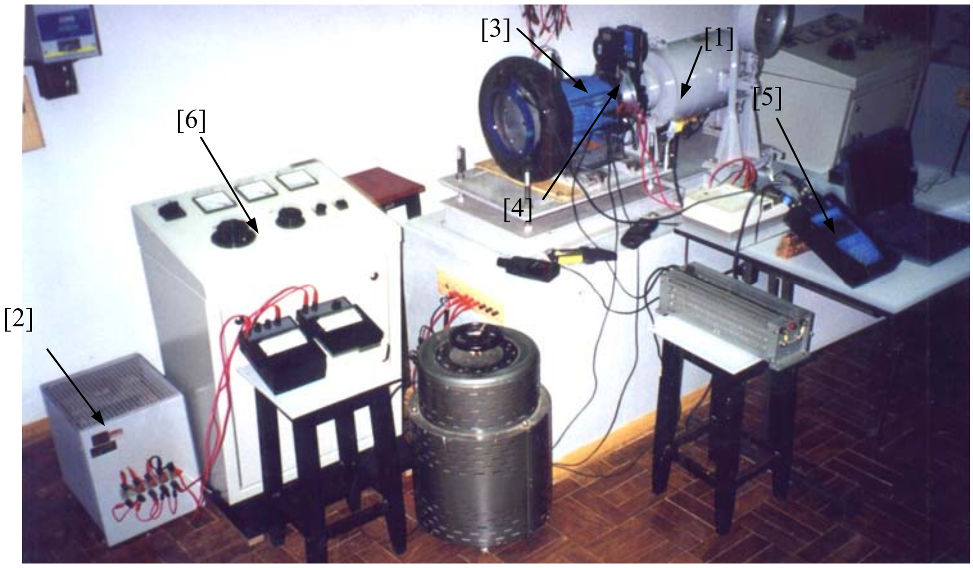


Figure 1 - Instrumented test desk.

The equipment UltraSepc 8000 [5] manufactured by the CSI (Computation System Incorporated) has been used for vibration spectra acquisition. It works collecting, memorizing and analyzing signals. Although a computer is not required to use the UltraSpec Analyze, data can be downloaded to UltraManager support software on a computer. The last option offers a more comfortable working environment for detailed analysis, technical reports writing and data bank generation.

The following equipment has been also used for parallel monitoring: high accuracy voltmeter "ENGRO- 600"; current digital clipper "*DAWER*- CM-600" and *Optho Tako* tachometer. The monitoring aim is to guarantee that the motor is working on the nominal load condition allowing that the simulated problems to become more visible in the spectra. The voltmeter measures give the voltage level information of the three phase motor supply. The motor load is adjusted altering the CC generator field current [6].

Though the UltraSpec 8000 specific firmware the system (motor plus generator plus load) has been laser alignment and precision balancing and it could be realized mechanical backlash. Dual visible laser beams and dual built-in inclinometers enable it to monitor the exact position of both shafts during rotation. Measurements are automatic and can be performed with or without cables. This method can determine shaft misalignment with less than one quarter turn of the shaft, making it ideal for application with restricted access. Thus vibration spectra could be obtained for the system under no failure condition (motor signature).

2. Spectra of broken bars through vibration analysis

Many technical articles have been addressed the importance of broken bars faults and have been developed differents techniques to diagnose them: Cho and Lang (1992), Elkasabgy and Eastham (1992), Elkasabgy *et al.* (1986), Kliman (1986), Kliman *et al.* (1989), Lowther and Silvester (1986), Penman and Stavrou (1996), Walliser and Landy (1994) and Williamson and Smith (1982).

Through the vibration analysis, among others, the studies for identification of imperfections in electric motors have an ample field to be explored. The anticipated detention of imperfections allows that the preventive maintenance is carried through during the coded stop of the machines. This period prevents a long period of stoping due a common imperfection in the electric motor increasing the availability of the system. The application of techniques of Artificial Intelligence in the detention of electric motors imperfections makes possible the accomplishment of the diagnosis in real time for a computer being able to present a minimum interaction with the user and, in some cases, without the aid of the specialists of the maintenance, Brito (2002).

An important cause of imperfection in the electric motors is related with the rotating bars. Therefore it is important to detect the imperfections in the first stages of development. The imperfection in an electric motor generally progresses in the following way: broken a rotating bar due the high tensions generated in the start-up; with the breaking the sparking appears causing additional heat and bending the rotor; despite the electric motor has been balanced the rotor tends to touch in the stator; adjacent bars receive more current and in this way the biggest mechanical tensions and heat are citizens; the blades of the rotor are get damaged leading the electric motor to the imperfection.

The difference between the rotation of an induction electric motor and its synchronous speed is known as slipping and can be calculated by the Eq. (1), where **s** is the slipping, **f** (Hz) is the frequency of the line **p** is the number of poles of the electric motor and **n** (Hz) is the rotation of the electric motor.

$$s = (2xf)/p - n \tag{1}$$

Torque, magnetic forces and frequency of the rotor are all modulated in the presence of a rotating bar. The change in the trajectory of the current due to broken bar produces harmonics fluxes that induce current in the rolling up of the stator, in the harmonic of the rotation frequency, with side bands in two times the frequency of slipping of the electric motor. Thus, the deterministic frequencies selected by the selective filter that correspond to the input net are: $\mathbf{1xf_r - 2xf_s}$, $\mathbf{1xf_r}$ and $\mathbf{1xf_r + 2xf_s}$, being the frequency of slipping ($\mathbf{f_s}$) equal to the slipping (**s**) times the frequency of line (**f**) of the line.

The accelerometer A0720GP, SN6714, accuracy de 0.1000 mV, has been used for vibration spectra acquisition. Hamming window of 3200 lines and 10 averages of samples has been used for a frequency width from 0 to 400 Hz and amplitude measured in speed (mm/s). The signals have been taken for the accelerometer at the vertical, horizontal and axial positions in both sides of the fan and the motor linking.

The vibration spectra for the normal working condition (motor signature), for the accelerometer at the vertical position in the motor linking side, are showed in Fig. 2. It can be seen from these spectra that the peaks have energy of 0,935 mm/s and 53,73 dB, respectively.

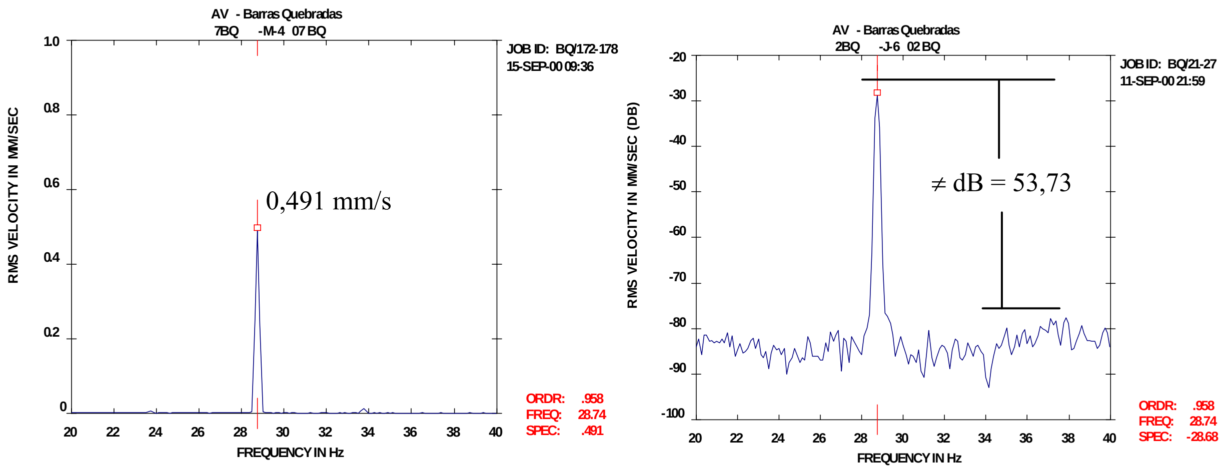


Figure 2 - Vibration spectra for the normal working condition.

The vibration spectra for 2, 3, 5 and 7 rotor broken bars, for the accelerometer at the vertical position in the motor linking side, are showed in Fig. 3 to 6, respectively. For the Fig. 3 the peaks have energy of 0,500 mm/s and 43,51 dB.

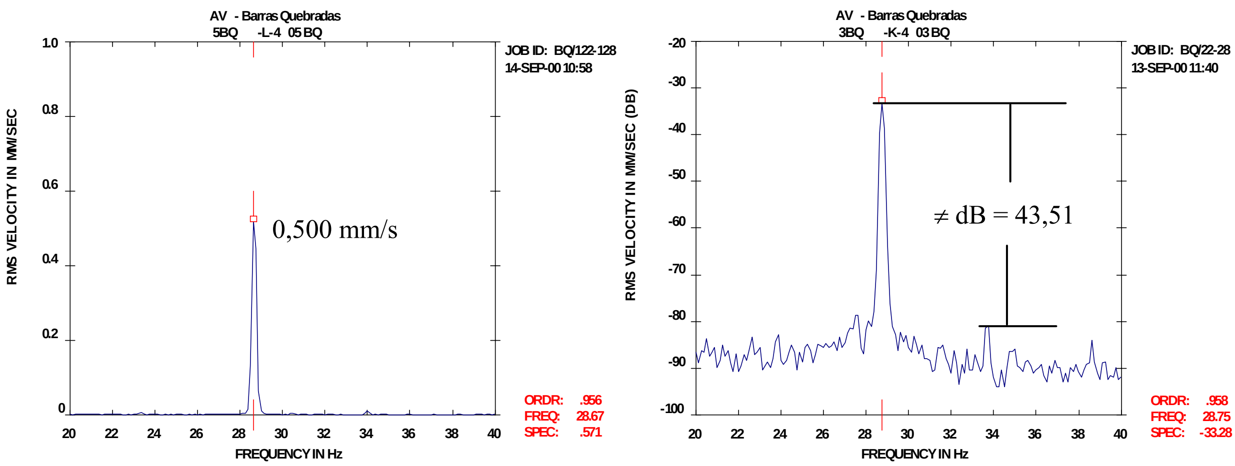


Figure 3 - Vibration spectra for 2 rotor broken bars.

For the Fig. 4 the peaks have energy of 0,550 mm/s and 46,43 dB. For the Fig. 5 the peaks have energy of 0,571 mm/s and 34,85 dB. Finally, for the Fig. 6 the peaks have energy of 0,491 mm/s and 31,91 dB. It can be seen from these spectra that the differences between the levels aren't greater. It's an important topic for the ANN's training and validation.

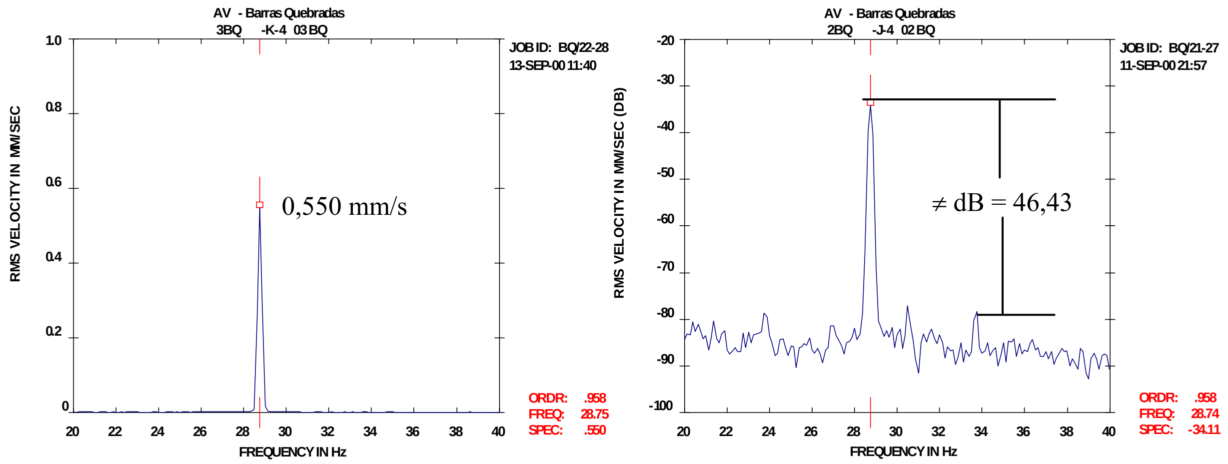


Figure 4 - Vibration spectra for 3 rotor broken bars.

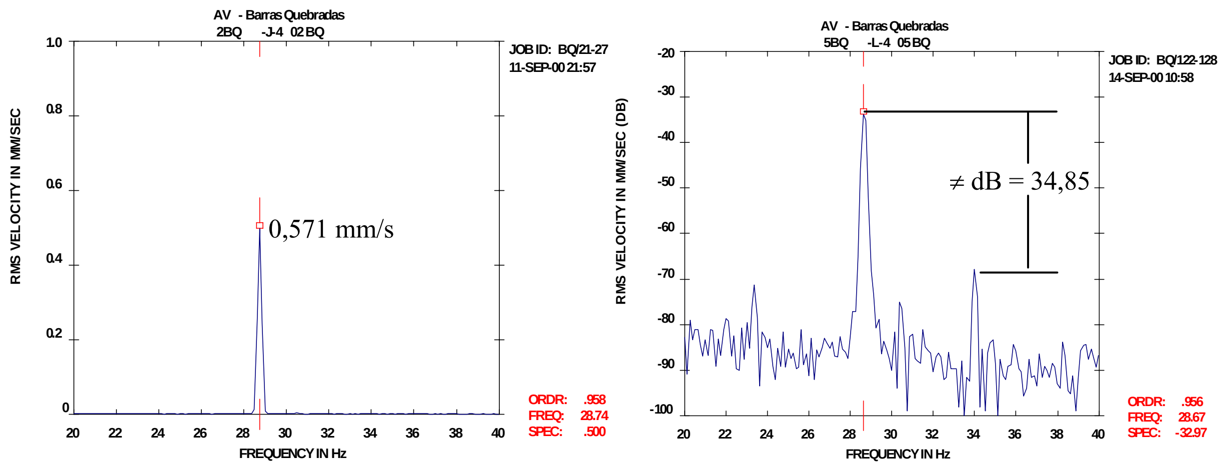


Figure 5 - Vibration spectra for 5 rotor broken bars.

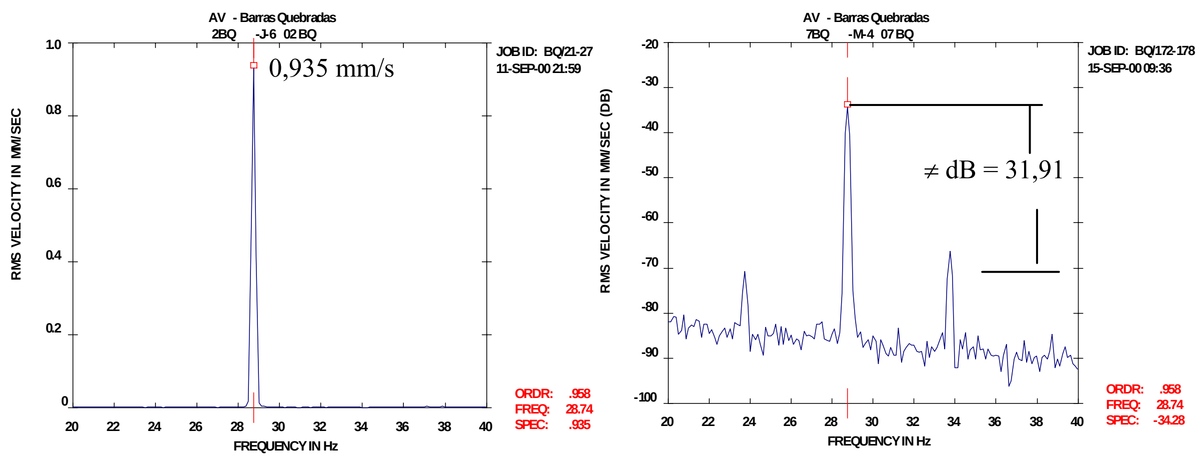


Figure 6 - Vibration spectra for 7 rotor broken bars.

3. Artificial Neural Networks

Many technical articles have been showed that the technology of artificial neural network has been used successfully in the detention of problems in electric motors: Chow and Yee (1990), Chow *et al.* (1991), Chow and Yee (1991), Goode and Chow (1993), Chow and Goode (1993), Chow *et al.* (1993), Chow (1994), Schoen *et al.* (1995), Chow (1997), Li *et al.* (1997), Filippetti *et al.* (1998) and Altug *et al.* (1999).

Artificial neural network (ANN) is a composed structure of simple processing units, distributed and parallel that has the intention to store knowledge by means of empirical mechanisms and to become it available for the use. It resembles itself to the human brain for the knowledge that is acquired by the network from the environment through learning mechanisms and by the "force" of the connections between the processing units they are minimized or maximized of form to better store the acquired knowledge, Haykin (2001).

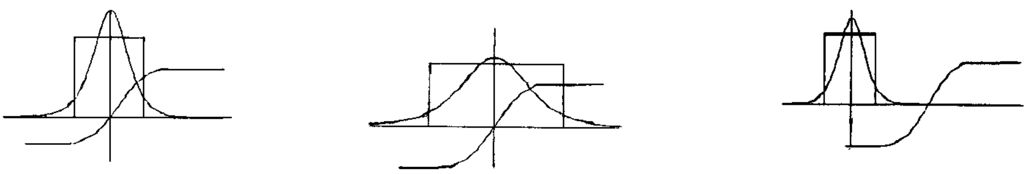
Artificial the neural networks are applied in many areas of the knowledge such as controlling of processes, analysis and processing of signals, classification of data, recognition of standards, analysis of images, medical diagnosis, etc. In the industrial area emphasize the used of neural networks in the prevention of shunting lines of processes in hybrid systems. The neural networks have been associating with the fuzzy logic techniques and specialists systems for detention of maintenance problems. Normally they are treated of problems with difficult mathematical quantification, inefficacious or even though impossible.

The main characteristic of the neural network is the capacity to learn and with this to improve its performance. They are trained with experimental data and the quality of these data exerts fort influence in the performance of the networks. The learning is the initial stage and consists of iterative processes of adjustment of the sinaptics weights and levels of bias binding. A set of initial conditions of a specific problem of optimization is presented to the neural net that will produce an exit. After to measure the distance between the real output and the desire output are realize the appropriate adjust in the weights of the connections in order to reduce this distance.

4. The Hybrid Paradigm

Agreed with Castro and Von Zuben (1998) the hybrid paradigm for initialization of the weights (INIT) can be interpreted as an intermediate method between the paradigms of the easiest way and the shortest way. Paradigm of easiest way supplies condition initial not necessarily next to solution excellent but that it is such that the training process can evolve more quickly in average, and more efficiently from the initial condition. Already the paradigm of the shortest way supplies possible the initial condition next to the excellent solution, still unknown.

In the hybrid paradigm the information contained in the training data is explored at the same time where it is tried to take in consideration the aspects of processing of signal of the net. The INIT searches to prevent that the intermediate units are saturated at the moment of the initialization of the weights defining adequately this set. The objective is to guarantee that the processing of the entrance data and the answers of the occult neurons are active initially in the linear part of the activation function as showed in Fig. 7a. The initial weights do not have to be defined of form to surpass the approximately linear region of the functions of activation for a subgroup of the training data as showed in Fig. 7b and for all the data of training as showed in Fig. 7c. It is evident that if the weights will be initialized of form that the internal reply of one or more units is of the linear region of the net for one or more standards of training the derivative of the activation function will be approximately zero. It becomes the process of training slower and subject to numerical problems, Castro and Von Zuben (1998).



(a) (b) (c)

Figure 7 - The hybrid paradigm for initialization of the weights (INIT). Source: Castro and Von Zuben (1998).

For the mathematical demonstration of the hybrid paradigm some considerations had been made. Consider **v** as the vector of weights for the intermediate layer, **X** the data set of entrance for training and **z** the exits (activation) of the intermediate units when **X** is presented to the network. The exit of intermediate units' **z** can be calculated, in matrix notation, for Eq. 2.

$$\mathbf{z} = \tanh(\mathbf{v}^T \mathbf{X}) \quad (2)$$

In case that the intermediate units are answering in the linear region of the activation function the Eq. (2) can be approached to the Eq. 3

$$\mathbf{z} \approx \mathbf{v}^T \mathbf{X} \quad (3)$$

The following question can be made: "Given the multiple inputs distributions what is the best combination of these inputs such that the combination of exit **z** is a normal distribution of average zero and small variance?"

The solution can be given by the following minimization according to the Eq. 4.

$$\min_{v \neq 0} \left\| v^T X - z \right\| \tag{4}$$

As it doesn't desire the trivial solution (**v = 0**) defines a distribution uniform with average zero and fixed variance for **z**. The solution for the Eq. (4) is given by Eq. 5.

$$\mathbf{v} = (\mathbf{X}\mathbf{X}^T)^{-1}\, \mathbf{z}^T \tag{5}$$

The following restriction is imposed in this procedure: matrix **X** has to have completed rank what it can easily be gotten simply manipulating the data of form entrance to eliminate redundancies.

Finally, the algorithm has to initialize the weights of the following layers and with small values since the exits of the first intermediate layer are characterized for a certain distribution around the linear band of the reply of the neurons that will be kept until the exit neurons. It's possible to initializes **Z** with a uniform distribution uniform in the interval of **[- 1,0; 1,0]** and **w = 0,1**.

5. Results

For each series of ten tests it has been separated randomly six series for training (total of 720 spectra) and four series for neural network validation (total of 480 spectra). Data inputs are generally compacted in order to reduce computational time and neural network's efficiency. In this way it has been implemented the selective filter in order to pick up only the deterministic frequencies of interest. This procedure reduced significantly the number of information to send to neural networks removing noises, redundancies and improving the quality of data training.

The great difficulty has been to identify those frequencies, which required an exhaustive analysis of all spectra. The frequencies of interest to identify broken bars are ($1 \text{ x } f_r \pm 2 \text{ x } f_s$ and $1 \text{ x } f_r$).

It has been built one artificial neural network to detect each of the five excitations (normal conditions, 2, 3, 5 and 7 broken bars) for each six positions of the sensors, in a total of thirty artificial neural networks. This procedure permits smaller architectures that are easier to train. A lot of tests have been done during the training in order to obtain the best architecture (3x3x3, 3x5x3, 4x4 and 7x4x7).

The artificial neural networks have been trained for the two methods of initialization of weights (random initialization and hybrid paradigm for initialization - INIT). The artificial neural networks have been trained with learning rule (backpropagation), activation function (hyperbolic tangent) and method of training (batch). During the training process it was possible the variation of the parameters of minimum error, rate of learning and constant of moment what it improved the convergence time.

During the test of validation each excitation has been passed in all thirty artificial neural networks and the condition of detected and undetected excitation has been considered. When one excitation has been presented for a specific artificial neural network the result has been considered detected for output values > 0,5 mm/s (1 mm/s is the ideal value) and ≤ 0.5 mm/s for the others (0 is the ideal value).

The first goal is to show the performance of the two methods of initialization of weights independent on the architectures used to training the artificial neural networks. In this way the artificial neural network output matrix for the normal condition and 2, 3, 5 and 7 broken bars are shown on Tab. (1). Agreed with Tab. 1 both methods of initialization of weights showed practically the same ratio of accuracy for diagnose fault (≅ 63%) even though any new training hasn't been performed.

Table 1. Ratio of accuracy of artificial neural network outputs for 2, 3, 5 and 7 broken bars.

Initialization Method	Ratio of accuracy		
	Diagnosed fault [%]	Not diagnosed fault [%]	Undefined fault [%]
INIT	63,33	10,63	26,04
Random	62,5	8,34	29,16

The second goal is to show the performance of the two methods of initialization of weights dependent on the architectures used to training the artificial neural networks. In this way the artificial neural network output matrix for 7 broken bars is shown on Tab. (2). Agreed with Tab. 2 the hybrid paradigm for initialization (INIT) method showed ratio of accuracy for diagnose fault (98,15%) better than random initialization method (90,75%).

The artificial neural network outputs of 7 broken bars presented better performance because their amplitudes in the deterministics frequencies have bigger levels than the others. For this reason the results from Tab. 2 are better than Tab.1.

Table 2. Ratio of accuracy of artificial neural network outputs for 7 broken bars.

Initialization Method	Ratio of accuracy		
	Diagnosed fault [%]	Not diagnosed fault [%]	Undefined fault [%]
INIT	98,15	0,00	1,85
Random	90,75	0,62	8,63

6. Conclusions

The inputs are one of the most important topics and have strongly influence on data convergence. If the base of data isn't well constructed the artificial neural network can present convergence problems. The tests procedures have been planned in detail in order to minimize ambiguity and mistakes during data acquisition. The data have been acquired randomly on the vertical, axial and horizontal directions, side of the fan and side of the motor linking. The vibration analysis has been chosen because it is a non invasive technology and has more information on the spectra belonging fault identification from mechanical and electrical sources. The domain of frequency has been chosen because it is easier to diagnose faults.

The rules have been classified in *diagnosed fault*, *not diagnosed fault* and *undefined fault*. The classification *diagnosed fault* means excitation has been diagnosed correctly. The classification *not diagnosed fault* means excitation has been wrongly diagnosed. Finally the classification *undefined fault* means that there aren't capable rules to identify the excitation.

With the gotten results a comparison between the used methods becomes very difficult. The results of Tab. 1 showed that the method of the hybrid paradigm does not possess great advantage related with the random method when applied in problems of different levels of broken bars.

However, the comparison does not imply in having better resulted and indeed the reason of them. The hybrid paradigm allows that only in the first iteration the hyperbolic tangent function answers in the linear region allowing a value adjusted for the local gradient. It allows not the saturation of the neurons of the net in the first iteration but it does not guarantee that such condition remains to the long one of the training being able to fall in poor local gradients what it does not result in excellent point.

On the other hand in the random method the amplitude of the deterministics frequencies (input of the ANN's) are small (less than 1) and the random initial set are in an appropriate interval [- 1, 1]. Thus the output of the network also is in the linear region what it becomes also favorable the random initialization.

7. Acknowledgements

The authors would like to express sincere appreciation to the FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais) and CNPq (Conselho Nacional de Pesquisa).

8. References

Altug, S., Chow, M-Y., 1999, "Fuzzy inference systems implemented on neural architectures for motor fault detection and diagnosis". IEEE Transaction on Industrial Electronics, Vol.46, No.6, pp. 1069-1079.

Brito, J. N., 2002, "Desenvolvimento de um Sistema Inteligente Híbrido para Diagnóstico de Falhas em Motores de Indução Trifásicos". Tese de Doutorado, Universidade Estadual de Campinas, Faculdade de Engenharia Mecânica.

Castro, L. N., Von Zuben, F. J., 1998, "Uma alternativa simples e robusta para inicialization de pesos em Redes Neurais Multicamadas". V Simpósio Brasileiro de Redes Neurais. Belo Horizonte, pp. 97-101.

Cho, K.R., Lang J.H., 1992, "Detection of broken rotor bars in induction motors using state and parameter estimation". IEEE Transaction of Industry Application, Vol.28, No.3, pp.702-708.

Chow, M-Y., 1997, "Methodologies of using neural networks and fuzzy logic technologies for motor incipient fault detection". Singapore: Word Scientific Publisher, 140p.

Chow, M-Y., 1994, "The advantages of machine fault detection using artificial neural network and fuzzy logic technologies". Proceedings of IEEE International Conference on Industrial Technology, pp.83-87.

Chow, M-Y, Goode, P. V., 1993, "Adaptation of a neural-fuzzy fault detection system". Proceedings of the 33rd Conference on Decision and Control.

Chow, M-Y, Mangum, P. M., Yee, S. O., 1991, "A neural networks approach to real-time condition monitoring of induction motor". IEEE Transaction on Industrial Electronics, Vol.38, No.6, pp.448-453.

Chow, M-Y, Sharpe, R. N., Hung, J. C., 1993, "Recognizing animal-caused faults in power distribuition systems using artificial neural networks". IEEE Transaction on Power Delivery, Vol.8, No.3, pp. 1268-1274.

Chow, M-Y, Yee, S.O., 1991, "Methodology for on-line incipient fault detection in single-phase squirrel-cage induction motors using artificial neural networks". IEEE Transaction on Industrial Electronics, Vol.6, No.3.

Chow, M-Y., Yee, S. O., 1990, "Real time application of artificial neural networks for incipient faul detection of induction machines". The Third International Conference of Industrial and Engeneering Applications of Artificial Intelligence and Expert Systems. Cahrleston, South Carolina.

Elkasabgy N. M., Eastham, A. R., 1992, "Detection of broken bars in the cage rotor on induction machine". IEEE Transaction of Industry Application, Vol.18, No.1, pp.165-170.

Elkasabgy N. M., Eastham, A. R., 1986, Dawson G. E, "The detection and effects of broken bars in cage rotor induction machines". In: Proceedings of the IEEE Workshop Electromagnetic Field Computation, Schenectady, New York, pp.G24-G28.

Filippetti, F., Franceschini, G., Tassoni, C., 1998, "Recent developments of induction motor drives fault diagnosis using AI techniques". IECON Proceedings Industrial Electronics Conference, pp. 1966-1973.

Goode, P. V., Chow, M-Y., 1993, "Neural/Fuzzy systems for incipient detection in induction motors". In: Proceedings. IECON'93, Honolulu, HI, Nov, pp.332-337.

Haykin, S., 2001, "Redes Neurais: Princípios e Prática". Trad. Paulo Martins Engel, Vol.1 No.2, Ed. Bookman, Porto Alegre, Brazil.

Kliman G. B., 1986, "The detection of faulted rotor bars in operating induction motors". In: Proceeding Internetional Conference of Elestric Machines (ICEM'86), Munich, Germany, pp. 500-502.

Kliman G.B., Koegl R. A., Stein J., Endicott R. D., Madden M. W., 1989, "Noninvasive detection of broken bars in operating induction motors". IEEE Transaction Energy Conversion, Vol.3, No.4, pp. 873-874.

Kolen, J. F., Pollack, J. B., 1990, "Back Propagation is Sensitive to Initial Conditions". Technical Report TR 90-JK-BPSIC.

Li, B., Goddu G., Chow, M-Y., 1997, "Knowledge based technique to enhance the performance of neural network based motor fault detectors". Proceedings of the IECON'97 23rd Internetional Conference on Industrial Eletronics, Control, and Instrumentation, Vol.3, pp.1113-1118.

Lowther D. A., Silvester P. P., 1986, "Computer-aided design in magnetics". New: Springer-Vertag.

Penman, J.., Stavrou, A., 1996, "Broken rotor bars: effects on the transient performance of induction machines". In: Proceeding of IEE Eletronic Power Apply, Vol.143, No.6, pp.449-457.

Schoen, R. R., Lin, B. K., Habetler, T. G., Schlag, J. H., Farag, S., 1995, "An unsupervised, on-line system for induction motor fault detection using stator current monitoring". IEEE Transactions on Industry Application. Vol.31, No.6, pp. 1280-1286.

Uliana, P. B., Seara, R. e Bermudez, J. C. M., 1998, "Inicialization Não Aleatória de Pesos de Redes Neurais Artificiais com Baixa Complexidade Computacional". Simpósio Brasileiro de Redes Neurais.

Walliser, R. F., Landy, C. F., 1994, "Determination of interbar current effects in the detection of broken rotor bars in squirrel cage induction motors". IEEE Transaction Energy Conversion, Vol.9, No.1, pp. 152-158.

Williamson, S., Smith, A. C., 1982, "Steady state analysis of 3-phase cage motors with rotor-bar and end-ring faults". Proceeding Institute of Elec. Eng., Vol.129, No.3, pp. 93-100.