Advanced forecasting and classification technique for condition monitoring of rotating machinery

Ilya Mokhov[1], Alexey Minin[2]

[1.] Saint-Petersburg State University, Applied Mathematics and Control Processes,
Ilya.Mokhov@siemens.com
[2.] Saint-Petersburg State University, Applied Computational Physics,
Alexey.Minin@siemens.com

Abstract. Prediction and classification of particular faults in rotating machinery, based on a given set of measurements, could significantly reduce the overall costs of maintenance and repair. Usually the vibration signal is sampled with a very high frequency due to its nature, thus it is quite difficult to do considerably long forecasting based on the methods, which are suitable for e.g. financial time series (where the sampling frequency is smaller). In this paper new forecasting and classification technique for particular vibration signal characteristics is proposed. Suggested approach allows creating a part of control system responsible for early fault detection, which could be used for preventive maintenance of industrial equipment. Presented approach can be extended to high frequency financial data for the prediction of "faults" on the market.

Keywords: fault analysis and prevention, artificial neural networks, artificial intelligence, rotating machinery, ball bearing failures, predictive monitoring**.**

1 Introduction

The faults in particular parts of industrial equipment could cause serious problems such as production losses, expensive repair procedures or even the personnel injures. Therefore the problem of fault analysis and prevention is very important. One of the main reasons of breakdowns of rotating machinery is the bearing failures. Plenty of papers, during the last decades, were devoted to the analysis of such kinds of faults by different methods of vibration analysis [11, 12, 13, 15]. The aim of this paper is to present the combination of forecasting and classification techniques, which could be used for the fault analysis and prevention.
In case the sampling rate is measured in kHz (for the dataset used in this paper sampling rate was equal to 40kHz), it is possible to estimate that prediction, achieved by the forecasting based on the pure time signal from the vibration sensors, will be only some microseconds ahead, which seems to be useless in terms of practical applications. Therefore in order to achieve applicable prediction for overcoming the problem described above should be introduced.

Let us consider the measurements obtained from rotating equipment containing ball bearings. Wide range of vibration analysis techniques could be used for analysis of the given measurements. The main idea of the traditional approaches is to analyze the peaks existence for particular frequencies and their multipliers (for more information regarding calculation of these frequencies and analysis methods [4, 5])

This paper will be devoted to the method which could be used for estimation of some particular defect frequency components evolution by using the artificial intelligence methods. The artificial neural networks, Neural Clouds [9] (further NC) and an advanced method of signal decomposition would be used in this research together with a trick of the particular fault related features extraction from Fourier spectrum.

It is out of the scope of this paper to describe the procedure of vibration data acquisition, for more information authors refer to [4]. In the following it would be assumed that a given time signal T was measured using vibration sensor for some ball bearing based system. The measurement strategy for efficient application of the presented method would be proposed in the following.

2 Preparation of inputs for the predictor

In this section data preprocessing and the preparation of the inputs for the forecasting and classification algorithms will be discussed. Following the work [6], it is a very important and challenging problem.

Let us divide given time sequence T into the set of equidistant sub sequences $\{t_i\}$ so that the time interval between them is equal to $\widehat{t}$. For the generalization of presented approach this procedure corresponds to the process of collecting the number of measurements periodically. It should be mentioned that dealing with a limited amount of measurements one should always remember about the tradeoff between how far one wants to forecast (it would be shown that it is related to the time interval between consequent measurements) and how many patterns one needs to make the prediction.

For every subsequence t_i the Fourier spectrum F_i is computed. Having the set of "consequent" Fourier spectra $\{F_i\}_{i=\overline{1,N}}$ the frequency dynamics, if any, could be observed. So that the monitoring of the important changes in signal frequency characteristic over the time can be done. This approach is related to the time frequency analysis (see fig. 1). Each of the obtained spectra could be used for analysis of the bearing conditions for the corresponding time interval. The idea here is to do forecasting in order to get the estimation of future spectral characteristic. In order to do it the particular frequency component changes over the sequence of spectra would be considered.

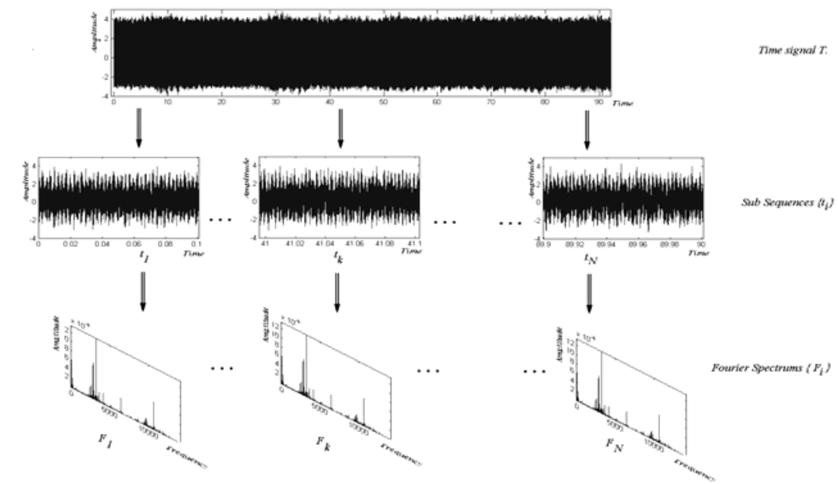

Fig. 1. Data preprocessing stage illustration

The problem is that the spectrum itself contains too many data and, moreover, not all these data are used for particular faults detection. Therefore it would be consistent to extract few features of the great importance regarding the particular type of fault from the overall frequency data set and to do forecasting only for them (see fig.2.). As a possible feature here an analogue of a crest factor measure C_j^k (see eq.1) can be used.

$$C_j^k = \frac{peak\ value}{RMS} = \frac{\max_i \{f_i\}}{\sqrt{\frac{\sum_{i=1}^{w_j} f_i^2}{w_j}}},\ j = \overline{1,M}\ and\ k = \overline{1,N} \tag{1}$$

where f_i are the frequencies from selected window and M is the number of features for every spectrum and N is the number of spectra. For $\forall j \in [1, M]$ values $\{C_j^k\}_{k=\overline{1,N}}$ form time series with a sampling rate inversely proportional to the time interval $\widehat{t}$ introduced above.

The overall spectrum feature extraction scheme could be described as a selection of important frequencies (e.g. the defect frequencies or the frequencies of particular interest), choosing an appropriate window around each of the frequencies, and then calculating the mentioned above crest factor analogues values for those windows (see fig. 3). Situation when the selected measure is close to 1 corresponds to the absence of peak within the window considered, while relatively high value corresponds to the existence of peak respectively.

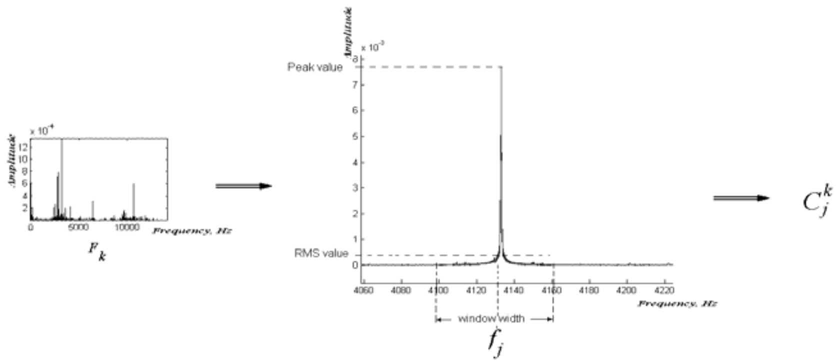

Fig. 2. Feature extraction idea.

Finally as an input for the predictor a set of the time series $\left\{C_j^k\right\}_{k=\overline{1,N}}$ is used, each of them corresponds to the selected particular frequency component j and has a sampling rate, defined by the time interval $\widehat{t}$ determined as a gap between two measurements. The aim of this process is to increase the time intervals between the data points, without loosing the important high frequency information.

One of the most important restrictions of the scheme presented is that all the measurements should be done within approximately similar conditions, such as rotation speed, external noise etc. Therefore one could expect that the defect growth will be observable within a selected frequency interval. Overcoming of this restriction is theoretically possible and could be considered as one of the further steps to extend the presented approach.

An example of the time series $\left\{C_j^k\right\}_{k=\overline{1,N}}$ is shown at the figure 3.

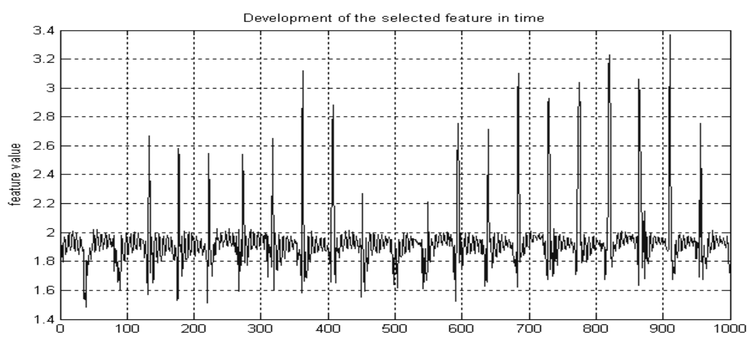

Fig. 3. Example of features time series we are going to predict

Then the Empirical Mode Decomposition [1, 2, 6] (further EMD) should be used in order to decompose initial signal into the set of orthogonal and non correlated time series (modes). The sum of the modes is equal to the initial signal (with an error of order 10^{-16}). Let us assume that the next state of the system will depend mainly on the current state of the system and depend also on few previous states with smaller weights (see fig.4). Therefore, the intrinsic mode functions [1, 2] at the current moment t and the current state of the system with a few lagged intrinsic mode functions (with smaller weights) should be used as inputs for the ANN. The result is shown in the figure 4.

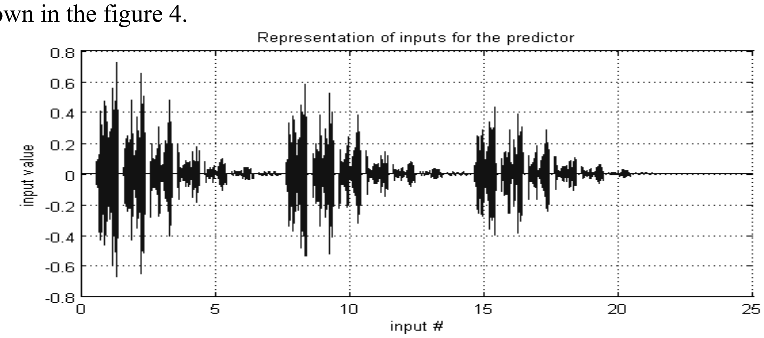

Fig. 4. Visualization of all inputs for ANN

As one can see from figure 4, some of the inputs (e.g. #6 and #7) are not of the great importance in comparison with the others. Nevertheless all of them have to be taken into account in order to avoid border effects of EMD method which could be rather strong [1]. Since the next state depends mainly on the current state (according to the assumption above) neural network have to be retrained to maintain the conditions of such specific process and to neglect border effects of EMD filter.

3 Forecasting technique and predictor architecture

Due to the fact that constructed time series are very specific, the architecture of predictor should be chosen properly. There are two main properties which the predictor should have. First one is good approximation capabilities; second one is presence of embedded memory, since time series represent some dynamical model. So far instead of simple feed-forward artificial neural network (further ANN), namely perceptron, which stands for the static mapping from the input to the output space (see fig.5. left hand side) it would be more convenient to use recurrent ANN, namely Elman ANN [7] (see fig.5. right hand side), since it implies the dynamic memory. Elman networks are also well known for their ability to extract features from time series. Moreover, recurrent networks are more robust in sense of over fitting, since this type of ANN corresponds to the so called infinite impulse response filter and require less parameter to describe system dynamics. Feed forward network corresponds to finite impulse response filter and need more parameters for mapping and therefore can lead to over fitting. Further the comparison of three architectures, namely time delay neural networks, linear regression and recurrent networks will be presented.

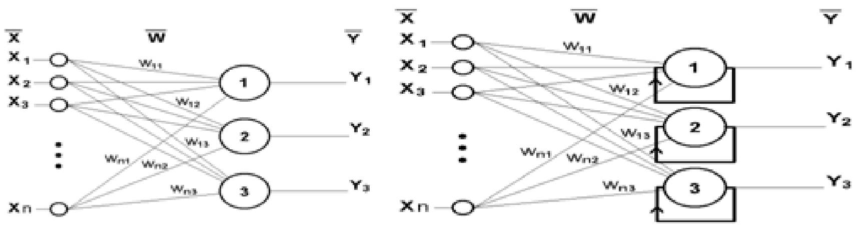

Fig. 5. $\{x_i\}$-input vector; $\{y_i\}$ – output vector; $\{w_{ij}\}$ – parameters (weights), Left hand side - Feed forward ANN; right hand side – RNN.

All parameters of the network, such as a number of layers and number of neurons were obtained from computational experiment. Number of layers was chosen to be 2. In each layer the number of neurons was chosen to be 40. Then the neural network committee which consisted of 10 ANN's was constructed. The result over the committee was simple averaging of the each network outputs. The length of the data was chosen to be 1000 time steps (1000 points). Training algorithm for Elman network was chosen to be BFGS algorithm [10], for time delay neural network training algorithm was chosen to be Levenberg-Marquardt [14]. Results of the forecast are presented in figure 6.

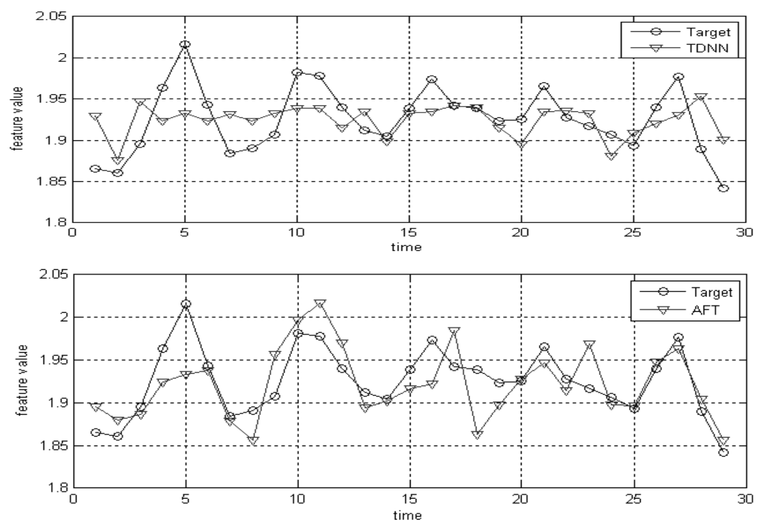

Fig. 6. One step forecast for the selected feature. The procedure was repeated 30 times to estimate statistical quality of prediction for one step

As it could be noticed from figure 6 (bottom picture), non delayed forecast for the frequency component development (see table 1) has been obtained. Below the comparison of linear regression and time delay neural network (further TDNN), with the proposed advanced forecasting technique based on RNN (further AFT) is given.

Table 1. Comparison of the quality of the forecast provided by different methods.

	r	R^2	MSE
Linear regression	0,30	0,06	0,03
TDNN	0,38	0,14	0,02
AFT	0,66	0,30	0,01

Table 1 gives a brief overview of the quality of the forecast for the following statistics: mean squared error (MSE), determination coefficient (R^2, see eq.2) and correlation coefficient (r).

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y}_i)}, \tag{2}$$

where $\hat{y}_i$ - is a forecast, y_i - is a true value and $\bar{y}_i$ - is an average over all training patterns. As one can see from table 1 AFT has a better correlation with the target than TDNN, and R^2 measure shows that determination of the forecast provided by AFT is better. The bigger determination means the lower probability of the forecast delay. Moreover using R^2 measure the optimal number of neurons in hidden layer could be estimated. The optimal number of neurons corresponds to the saturation of R^2 measure, so that the increasing of the hidden layer does not improve the forecast.

The forecast was performed in a same manner for all features being extracted from Fourier spectrum, so that for all selected signal characteristics the future values were obtained. By analyzing the obtained values future machine state could be estimated. To make the procedure of classifying the machine condition as normal or abnormal automatic, the usage of a novel classification technique would be proposed in next section.

4 Classification technique and overall scheme

To make process fully automatic and suitable for implementation of fault prevention system it was decided to add mechanism for classification. For this purpose the novel technique which is called “Neural clouds” [9] was chosen. Since the details of NC are not published at the moment, its main ideas will be briefly discussed in the following.

The concept which stands behind the NC term consists of creating an efficient data encapsulation mechanism for the so called one-side classifier, using the advanced clustering algorithm and extended Radial Basis functions network approach. The basic idea of the one-side classification usage in field of condition monitoring, fault analysis and prevention is that the real industrial data, which could be collected on the plant, usually correspond to the normal conditions, while the bad data collection is expensive, and fault modeling is not always available. The system is trained on a given dataset which corresponds to the normal operating conditions of the plant and stands for the detection of the fluctuations from these conditions.

The simple example of classification is shown in figure 7. Here the data points are displayed with circles and confidence levels are displayed with lines. The plateau on the fig.7 right hand side corresponds to the conditions which suppose to be normal with probability equal to 1 and the slopes shows how the probability decreases down to 0.

After obtaining a new forecast one just “drops” or project it on the classifier (see fig.7.) and chooses the nearest confidence level. Then one should decide whether forecasted state induces alarm or not, using the simple formula: $P_f = |1 - Conf|$, where P_f is probability of failure and $Conf$ is confidence level.

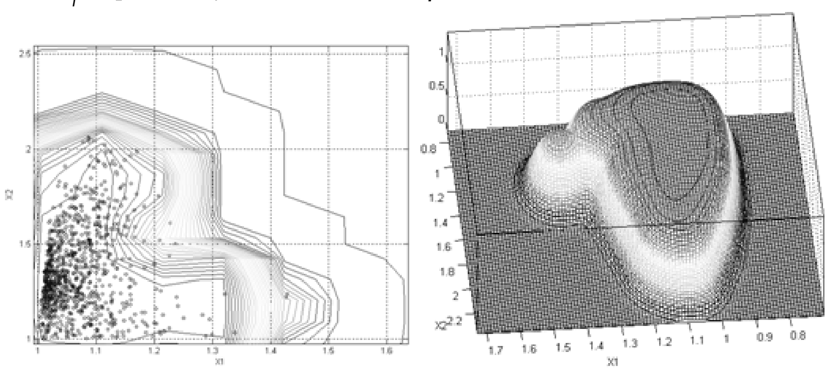

Fig. 7. Left hand side: Classification in 2D space for 2 selected features (normalized density), right hand side: classifier in 3D space (normalized density).

The overall scheme of the system is like following: first extract features from Fourier spectrum, then comes AFT to provide forecast of particular Fourier spectrum features and NC to recognize whether the predicted state is normal or abnormal. After the system is trained in back testing mode, AFT has to be retrained after each step, while NC remains the same if the normal conditions do not change. Overall scheme is shown in figure 8.

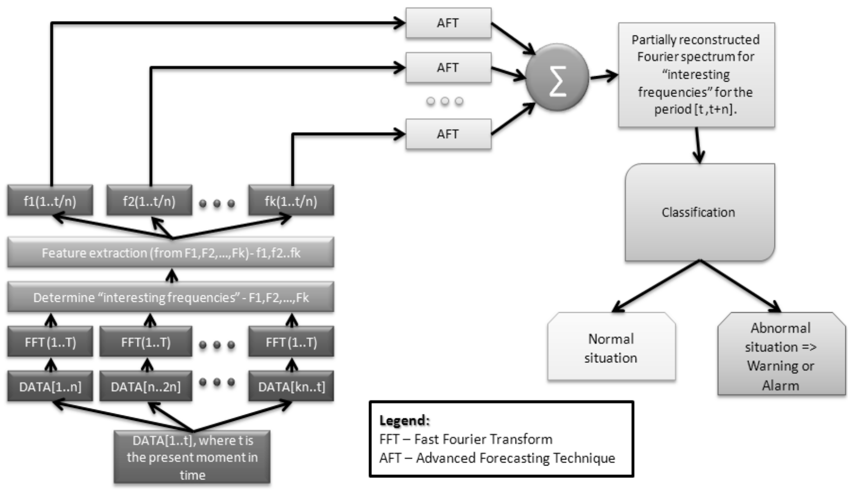

Fig. 8. Overall scheme of the system

5 Conclusions and outlook

The most of the neural network applications in the field of vibration diagnostic are devoted to the classification problem. In the presented paper authors made an attempt to apply it to the forecasting problem. The approach presented above could be considered as an additional part of the vibration analysis and monitoring system. System offers the experts a possibility to analyze the evolution of the particular defect, introduced by the bearings components and to perform predictive monitoring of the future state of the system. Additional classification stage makes the abnormal behavior detection process automatic. Together with a suitable measurement strategy it could be used for a considerably long term prediction of frequency characteristics of the vibration signal and therefore the prediction of the system conditions.

According to the set of experiments it was found that the proposed technique could be used for the prediction of the introduced frequency features.

As a possible extension of the suggested approach a forecasting for the additional values could be considered (e.g. forecasting enveloping spectrum features, some additional measurements and calculated statistical values (for instance kurtosis, skewness)). This will allow covering the forecasting of the main characteristics of vibration signal used for the analysis of the equipment conditions. Taking into account changing rotation speed one should consider as the thing of a great importance simplifying the data acquisition process.

The extension of the forecasting horizon to middle range could be considered on the basis of the paper [8].

Acknowledgments. The work was done with support of OOO “Siemens” CT, namely Bernhard Lang (Bernhard.lang@siemens.com), “Fault analysis and prevention” and with support of Kuperin Yuri (Yuri.Kuperin@gmail.com), Saint-Petersburg State University.

References

1. Gabriel Rilling, Patrick Flandrin and Paulo Gonëcalv`es, On empirical mode decomposition and its algorithms, IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing NSIP-03, Grado (I)
2. Norden E. Huang, Zheng Shen, Steven R. Long, Manli C. Wu, Hsing H. Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung and Henry H. Liu, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Mathematical, Physical and Engineering Sciences, Vol. 454, No. 1971 (Mar. 8, 1998), pp. 903-995
3. Ivan Magrin-Chagnolleau and Richard G. Baraniuk, Empirical mode decomposition based on it’s time-frequency attributes, Proceedings of the 69th SEG Meeting, Rice University, Houston, Texas.
4. N.A. Barkova, Introduction to Vibration Diagnostics of Rotating Machinery and Equipment, SPbSMTU 2003 (in Russian).
5. A.V. Barkov, N.A. Barkova, A.U. Azovcev Monitoring and Diagnostics of Rotor Machines by Vibration, SPbSMTU 2000 (in Russian)
6. A. Minin, Y. Kuperin, B. Lang, Increasing the quality of neural forecasts with the help of Empirical Mode Decomposition, Proceedings of the International Scientific Conferences “Intelligent systems” (AIS’07) and “Intelligent CAD’s” (CAD-2007), Physmathlit, vol.4, pp. 68-74
7. Narendra K.S., Parthasarathy K. Identification and control of dynamical systems using neural networks. IEEE Trans. Neural Networks, vol.1, pp.4-27, Mar. 1990.
8. Kevin Ludd, Michael Small Towards long-term predictions, Physica D 136 (2000) pp.31-44.
9. B. Lang, I. Mokhov Neural Clouds for monitoring of complex plant conditions (to appear)
10. Broyden C. G. Journal of the Institute of Mathematics and Its Applications 6 (1970), pp. 76-90
11. Javier Sanza, Ricardo Pererab, Consuelo Huertab, Fault diagnosis of rotating machinery based on auto-associative neural networks and wavelet transforms, Journal of Sound and Vibration 302 (2007), pp. 981–999
12. A.J. Hoffman, N.T. van der Merwe, The application of neural networks to vibration diagnostics for multiple fault conditions, Computer Standards & Interfaces 24 (2002), pp.139–149
13. Menderes Kalkat, Sahin Yildirim, Ibrahim Uzmay, Design of artificial neural networks for rotor dynamics analysis of rotating machine systems, Mechatronics 15 (2005), pp. 573–588
14. Hagan, M.T., M. Menhaj, Training feed forward networks with the Marquardt algorithm, IEEE Transactions on neural networks, vol.5, no.6, pp. 989-993.
15. Ian Howard, A review of rolling element bearing vibration “Detection, diagnosis and prognosis”, DSTO Aeronautical and Maritime Research Laboratory (1994).